

Portland State University
PDXScholar

Dissertations and Theses

Dissertations and Theses

Spring 6-6-2018

A Study of Fragmentation and Spontaneous Covalent Self-
Assembly of the *Azaorcus* Ribozyme from Multiple Small Inactive
RNA Fragments

Tharuka Sewwandi Jayathilaka
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

 Part of the [Chemistry Commons](#)

Recommended Citation

Jayathilaka, Tharuka Sewwandi, "A Study of Fragmentation and Spontaneous Covalent Self-Assembly of the *Azaorcus* Ribozyme from Multiple Small Inactive RNA Fragments" (2018). *Dissertations and Theses*. Paper 4397.

[10.15760/etd.6281](https://pdxscholar.library.pdx.edu/open_access_etds/10.15760/etd.6281)

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

A Study of Fragmentation and Spontaneous Covalent Self-Assembly of the

Azoarcus Ribozyme from Multiple Small Inactive RNA Fragments

by

Tharuka Sewwandi Jayathilaka

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Chemistry

Dissertation Committee:

Niles Lehman, Chair

Michael Bartlett

Albert S. Benight

Andrew Rice

Reuben H. Simoyi

Portland State University
2018

© 2018 Tharuka Sewwandi Jayathilaka

Abstract

The question about the origins of life often appears as a difficult question to answer. A more reliable candidate molecule for the chemical origins of life would be a molecule that is capable of making copies of itself from simple precursors. With the finding of the catalytic activities of RNA molecules by Thomas Cech and Sid Altman in late 1980s, the term ribozyme was introduced to define an RNA molecule with catalytic activity. The RNA World is a conceptual period in the early stages in the development of life because RNA simultaneously possesses evolvability and catalytic function. An RNA molecule that could evolve in such a fashion is likely to have been one of the Earth's first life forms. The most important problem facing the RNA World is the difficulty of prebiotic synthesis of RNA. Different prebiotic environments could provide the right reaction conditions for synthesis of catalytically active RNA molecules. Most importantly, these environments can support new ways to assemble monomers into polymers.

In order to understand and demonstrate how small inactive RNA oligomers can self-assemble into an autocatalytic ribozyme molecule, here I have used the *Azoarcus* Group I intron. First, the fragmentation and the self-assembly reactions were done using the natural IGS-tag combination of the *Azoarcus* group I intron, which is GUG/CAU. The main purpose was to experimentally demonstrate that the ribozyme can be broken down into five or more shorter RNA fragments and these fragments can self-assemble into a catalytically active covalent full-length molecule. Then, with the successful demonstration that five

inactive RNA fragments can self-assemble, the next step to test the other possible breaking locations with mutated IGS/tag combinations. A new IGS/tag pair GCG-CGU also successfully demonstrated the five-piece self-assembly reaction. Finally, we tested these reactions in different Mg^{2+} concentrations to optimize the self-assembly reactions. By focusing not only one single reaction but on a collection of different sequence requirement combinations and with the development of evaporation and subsequent rehydration by spinning down methods this study successfully illustrates that covalent self-assembly from inactive RNA oligomers is possible. Therefore, this thesis work focuses on a more broader aspect of intermolecular interactions in the study of the RNA World, as illustrated in following chapters.

Dedication

To my mother Lalani Nanayakkara and to the memory of my father Ranjith Jayathilaka.

To my loving sons, Sahas Alahakoon and the one who is soon to be born. You both have made me a stronger person I could have ever imagined.

Acknowledgements

I would first and foremost like to thank my advisor, Dr. Niles Lehman for his incredibly supportive and wise mentorship. His guidance through this scientific journey has truly transformed my ability to consider and question the world around me.

Then I would like to thank my committee members, Drs. Albert S. Benight, Michael Bartlett, Reuben Simoyi, and Andrew Rice. A huge thank you for your time commitment, suggestions, corrections, and conversations.

Last, but not least to my co-workers, fellow graduate and undergraduate students in Lehman lab, and my colleagues from the Department of Chemistry, who have given me more encouragement and support than I will ever have time to fully thank.

Finally, to my family, my mom and dad, I would not be able to complete this work without your support and encouragement. To my husband and my babies gave me the tremendous courage and inspiration for me to achieve my dreams and goals.

Table of Contents

Abstract	i
Dedication.....	iii
Acknowledgements	iv
List of Tables.....	vi
List of Figures.....	vii-viii
Chapter 1	
Introduction.....	1-13
Chapter 2	
Self-Assembly of the <i>Azoarcus</i> Group I Intron from Five Small RNA Fragments.....	14-47
Chapter 3	
RNA Self-Assembly Experiments with Multiple Different IGS-Tag Combinations	48-78
Chapter 4	
Analysis of Effect of Evaporation and Re-hydration on RNA Self-Assembly.....	79-101
Chapter 5	
Techniques for Establishing an RNase- and PCR Contamination-free Research Laboratory.....	102-115
References.....	116-123

List of Tables

Table 1: Nucleotide sequences at the W-X, X-Y, Y-Z1, and Z1-Z2 junctions.....	40
Table 2: RNA oligomers used in self-assembly reactions using GUG-CAU combination.....	47
Table 3: Different approach to test self-assembly reactions with other IGS-tag combinations	52
Table 4: Nucleotide sequences at the junctions in six-piece self-assembly.....	74
Table 5: The change of total reaction volume, RNA and MgCl ₂ concentrations after six hours of incubation.....	96
Table 6: The change of total reaction volume, RNA and MgCl ₂ concentrations after eight hours of incubation.....	97
Table 7: The change of total reaction volume, RNA and MgCl ₂ concentrations after nine hours of incubation.....	98
Table 8: Percent molar yields of five-piece self-assembly, depending on total reaction time and time of immediate centrifugation.....	98

List of Figures

Figure 1: Breaking strategy of the <i>Azoarcus</i> group I intron.....	12
Figure 2: The two-step process of group I intron self-splicing <i>in vivo</i>	21
Figure 3: The secondary structure of the <i>Azoarcus</i> group intron with all five fragments: W, X, Y, Z1, and Z2.....	22
Figure 4: The two possible mechanisms of RNA-catalyzed recombination	23
Figure 5: Two-piece self-assembly with X fragment: WX1+X2YZ	26
Figure 6: Three-piece self-assembly with X fragment: W+X1+X2YZ.....	27
Figure 7: Four-piece self-assembly with X fragment: W+X1+X2+YZ.....	28
Figure 8: Two-piece self-assembly with Y fragment: WXY1+Y2Z.....	30
Figure 9: Three-piece self-assembly with Y fragment: W+XY1+Y2Z.....	31
Figure 10: Two-piece self-assembly with Z fragment: WXYZ1+Z2.....	34
Figure 11: Three-piece self-assembly with Z fragment: W+XYZ1+Z2.....	35
Figure 12: Four-piece self-assembly with Z fragment: W+X+YZ1+Z2.....	36
Figure 13: Five-piece self-assembly with Z fragment: W+X+Y+Z1+Z2.....	37
Figure 14: Self-assembly reactions with GCG-CGU combination.....	53
Figure 15: Two-piece self-assembly reactions with GCC-GGU combination.....	56
Figure 16: Three and four-piece self-assembly reactions with GCC-GGU combination.....	57
Figure 17: Self-assembly reactions with GCA-UGU combination.....	59
Figure 18: Self-assembly reactions with GUU-AAU combination.....	61
Figure 19: Self-assembly reactions with GUC-GAU combination.....	63
Figure 20: Self-assembly reactions with GUA-UAU combination.....	65
Figure 21: Self-assembly reactions with GGG-CCU combination.....	67
Figure 22: Self-assembly reactions with GGC-GCU, GAU-AUU, and GAG-CUU combination.....	69

Figure 23: Combined self-assembly reaction; six-piece self-assembly.....	72
Figure 24: Illustration of testing the effect of evaporation and rehydration.....	89
Figure 25: Self-assembly reactions using a heat block and thermal-cycler.....	90
Figure 26: Effect of the length of the evaporation time on five-piece self-assembly.....	91
Figure 27: Effect of different Mg^{2+} concentration on five-piece self-assembly.....	92
Figure 28: Effect of 50 mM and 100 mM Mg^{2+} concentration on five-piece self-assembly.....	93
Figure 29: Effect of 25 mM and 100 mM Mg^{2+} concentration on five-piece self-assembly.....	94
Figure 30: Effect of using different total reaction volumes on five-piece self-assembly.....	95
Figure 31: Illustration of different layers of a DNA and RNA research laboratory.....	106
Figure 32: Illustration of the work flow of a PCR laboratory.....	110

Chapter 1

Introduction

The study of how life originated on the Earth has become one of the most interesting and yet difficult challenges, because it happened some 4 billion years ago, and we do not have the detailed knowledge of the process that led to the emergence of life on early Earth. No direct evidence of its occurrence has survived that long. As well as it is not possible to assign a chronological order to the appearance of life. In fact, life remains only vaguely defined. As Gerald Joyce stated in the NASA's working definition for life, it is "a self-sustaining chemical system capable of undergoing Darwinian evolution" (Joyce, 2002). An alternative definition to this is that the first living entities "were systems that are capable of undergoing Darwinian evolution, whose self-sustaining properties depend on the availability of organic molecules already present in the primitive environment" (Lazcano, 2007).

One of the essential principles of beginning of life on the Earth is that the formation and the replication of the first main living system depended primarily on prebiotically synthesized organic molecules (Delaye and Lazcano, 2005). This is supported by many successful prebiotic simulation experiments in the past century. Some of the most famous experiments of these are synthesis of urea in 1828 by Wöhler (Wöhler, 1828), synthesis of sugar molecules by Butlerow in 1861 (Butlerow, 1861), and synthesis of amino acids by

simulating prebiotic atmospheric conditions inside a glass apparatus by Stanley Miller (Miller, 1953).

A few years after the Miller experiment, Joan Oró, who studied the synthesis of amino acids from an aqueous solution of HCN and NH₃, and reported the abiotic formation of adenine (Oró, 1960). With the discovery of the X-ray diffraction structure of the double stranded DNA molecule in 1953 by Watson and Crick, the importance of the nucleic acids arose in prebiotic studies (Watson and Crick, 1953). In modern day biology nucleic acids and proteins play a major role in central biological functions.

Once these abiotic monomers have started to accumulate in early prebiotic environments, these became the raw building blocks for the synthesis of larger polymers. Eventually the formation of such polymers led to pathway/s to the beginning of the life as we know today. Previous laboratory experiments demonstrate that presence of mineral surfaces can induce the formation of oligomers up to 55 nucleotides through polymerization (Ferris et al., 1996). Absorption of molecules onto mineral surfaces like montmorillonite can increase concentration of the reactants and favors polymerization of the activated monomers. When these polymerized molecules became more complex, some of them can start folding into certain favorable configurations and interact with other complex molecules. The diversity of these complex molecules and the reactions among them portrays the plausibility of how chemistry can spontaneously build complexity.

There is little doubt that nucleic acids were the first genetic material, and that DNA and RNA are the practicable genetic materials for directed *in vitro* evolution. The concept of RNA as a primordial molecule is very old. The ideas behind the hypothesis of an RNA world originated in the late 1960s. In 1962 Alex Rich proposed that RNA might be able to support both genetics by being copied and phenotype by having catalytic activity (Rich, 1962). Since then, many suggested that if an RNA molecule could perform most of the catalytic functions that are currently accomplished by proteins, then it was possible to have an autonomous RNA organism during the early stages of formation of life on the Earth (Woese, 1967; Crick, 1968). With the discovery of the ribozyme, (Kruger et al., 1982) many research studies showed that RNA molecules catalyze a diverse array of chemical reactions. This shed the light to many significant discoveries using catalytically active RNA molecules and suggestions regarding an “RNA World”, a term introduced by Walter Gilbert in 1986 (Gilbert, 1986). The most explicit evidence for an RNA world is perhaps that the RNA in the ribosome catalyzes peptide-bond formation (Nissen et al., 2000). But it is believed that a pure RNA world lasted only for a relatively small period of Earth’s history, mainly due to the lower stabilities and catalytic ranges of primitive RNA molecules compared to proteins. Nevertheless, there is evidence of an extensive catalytic repertoire of ribozymes.

Most of the known ribozymes catalyze phosphoryl transfer reactions, in which ribozymes carry out site-specific cleavage of RNA via nucleophilic attack. RNA, as a catalytic macromolecule, has a certain number of resources to perform its catalytic function. An

RNA molecule is made up of four nucleobases, which are connected by a ribose-phosphate backbone. These nucleobases can participate directly in the catalytic process using their nitrogen atoms in the purine or pyrimidine rings. They act as hydrogen bond donors and help to stabilize the transition state. Metal ions also play an important role in folding of RNA molecule, by helping to establish and maintain the active geometry of the ribozyme (Woodson, 2005). X-ray crystal structures have provided much valuable insight into ribozyme mechanisms (Scott et al., 1996).

Group I introns, one of the 14 classes of naturally-occurring ribozymes, can be found in a wide range of organisms including, fungi, algae, and many other unicellular eukaryotes. The prevalence of the group I introns across many taxa could be a result of their ability to self-splice from an RNA transcript, a mechanism for which is known in detail (Price and Cech 1988). Several structural studies have provided details on different stages of the splicing reaction of group I intron ribozymes. Self-splicing in Group I introns occurs through two transesterification reactions (exchanges of phosphate esters that leave the total number of phosphodiester bonds unchanged). In the first transesterification, the 5' splice site is cleaved, as guanosine is added to the 5' end of the intron; guanosine, GMP, and GTP have similar activities as substrates. Guanosine acts as a nucleophile in an in-line, S_N2 reaction. In the second transesterification step, the 3' splice site is cleaved as the exons are joined.

There are three functionally important metal-ion binding sites available in these group I introns (Shan et al., 1999). In the first transesterification reaction, first metal ion, M_A is bound to a non-bridging oxygen atom of the scissile phosphate and the 3' hydroxyl of the 5' splice site. Then the second metal ion, M_C is bound to the same non-bridging oxygen atom and the 2' hydroxyl of the attacking guanosine, while the third metal ion, M_B is bound to the 3' hydroxyl of the attacking guanosine. Metal ions bind at the active sites of all the group I ribozyme structures (Guo et al., 2004) (Adams et al., 2004).

In 1992, Reinhold-Hurek and Shub discovered a group I intron after third anticodon nucleotide in the tRNA^{Ile} from the purple bacterium *Azoarcus* (Reinhold-Hurek and Shub, 1992). The crystal structure of the *Azoarcus* ribozyme at 3.1 Å resolution was resolved with RNA fragments representing both 5' and 3' exons and poised at the step before the second transesterification reaction that will join the exons (Adams et al, 2004). The architecture core of the ribozyme is conserved very well. The overall structure consists of three coaxially stacked helical associations comprising the P1 and P2 helices, the P4-P6 helices, and the P9, P7, P3 and P8 helices. These are linked by several tertiary interactions. The active center of the *Azoarcus* ribozyme has a very high density of phosphate groups, which coordinate two metal ions corresponding to M_A and M_C ; these have been assigned to Mg^{2+} and K^+ ions, respectively. The Mg^{2+} ion is maximally coordinated to phosphate oxygen atoms, with five out of six inner-sphere ligands exchanged (Golden et al., 2005). Importantly, the *Azoarcus* intron can retain its catalytic activity under extreme reaction conditions, such as higher temperatures ranging from 70°C-80°C, higher urea

concentrations up to 7M, and at higher pH values such as 10.1. This may be due to its high G-C content (60%) and two stable GAAA tetraloops. Moreover, in the first step of splicing, this intron has a $k_{\text{cat}}/K_{\text{m}}$ value that is greater than most group I introns. Unlike in the case of *Tetrahymena* ribozyme, for the *Azoarcus* ribozyme, substrate release is not rate limiting (Tanner and Cech, 1996). This makes the *Azoarcus* ribozyme a valuable and versatile model system for group I intron related prebiotic studies.

The main principle of the RNA World is that an RNA molecule could make a copy (or more copies) of itself. This idea of RNA replication has been studied in the laboratories using several different approaches. The development of networks of molecules that catalyze the formation of other specific sequences where the products are autocatalytic is one of the main approaches (Higgs and Lehman, 2015). The term “self-replication” not only refers to the production of additional copies, it also refers to ability to make more copies that behave in a similar manner. This situation would exhibit the autocatalytic behavior with the ability for exponential growth (Kim and Joyce, 2004). A hypothetical RNA replicase can act both as a template and as an RNA polymerase that can replicate its own sequence. For example, the class I ligase was evolved from starting pool of more than 10^{15} random RNA sequences (Bartel and Szostak, 1993). Because of its potential as a primordial RNA replicase, the class I ligase represents an obvious goal for RNA-directed recombination from smaller RNA pieces. Even though there is no direct evidence of presence of such RNA replicase molecule, several previously performed laboratory

experiments show that RNA is capable of performing catalytic activities. However, these ligase molecules have two significant challenges. First, they are moderately sized molecules of 100-160 nt in length. Second, they are structurally very complex molecules. Still, one of the important questions for the beginning of an RNA World regards how the first template-dependent RNA polymerase ribozyme emerged (Briones and Manrubia, 2009). Orgel suggested that “one must postulate that a library of RNA strands with different sequences formed spontaneously on the primitive Earth and that this family of sequences included catalysts able to support self-replication of RNA” (Orgel, 2004).

Recombination is the exchange of genetic information between two sources. Recombination can happen when two homologous regions combine with each other or crossover between two regions. This will result in both longer and shorter recombinant products. However, in non-recombinant template-directed polymerization, polymer length increases by addition of one or more nucleotides in several rounds of monomer addition until the desired polymer length is reached. One of the main advantages of recombination is that it does not rely on the use of high-energy leaving groups. Instead, exchange of oligomers conserves the enthalpy stored in phosphodiester bonds. Recombination can be seen in naturally occurring ribozymes. For example, any derivative of guanosine will drive the Group I ribozyme catalytic reaction. With the discovery of *Tetrahymena* ribozyme, one could demonstrate its ability to catalyze *in vitro* production of RNA oligomers up to 30 nucleotides in length, starting from 5-mers (Bartel and Szostak, 1993).

The presence of pools of different RNA molecules for further reactions under the Earth's early prebiotic environmental conditions was affected and enhanced by several factors including temperature fluctuations, freezing/melting cycles, different ionic strengths, metal ion concentration, evaporation processes, and watering cycles in tidal regions. Most importantly wet-dry cycles provided a favorable condition in the ligase activity of the hairpin ribozyme (Kazakov et al., 2006) and the encapsulation of the macromolecules by membrane bounded vesicles (Deamer, 1998) (Briones and Manrubia, 2009).

In order to study condensation products of mononucleotides and hydrolysis of their polymers, a study using 5'-mononucleotides was done and observed that the mononucleotides were able to organize and produce oligomers within a multilamellar lipid matrix using anhydrous phase of hydration–dehydration (HD) cycles (DeGuzman, 2014). Previous studies show that freezing can stimulate the self-ligation of linear forms of the hairpin ribozyme which contains 2',3'-cyclic phosphate and 5'-OH termini. Under freezing conditions, the cleavage-ligation equilibrium strongly favors the ligation. Moreover, evaporation of the aqueous solvent can also support the hairpin ribozyme ligation activity and the lowered water activity can enhance the ligation of the ribozyme. This concept might have relevance to the prebiotic evolution of the other ribozymes (Kazakov et al., 2006).

RNA-like polymers can be synthesized non-enzymatically from mononucleotides in lipid environments, where RNA can be synthesized at elevated temperature ranges in fluctuating environments (Rajmani et al., 2008). In such environments, cycles of wetting and drying have sufficient free energy to start off a mixture of simple molecules towards a complex molecular system. The concentration of the dilute organic solutes increased during the drying process (Rajmani et al., 2008). Several studies showed the possibility for the synthesis of RNA under hydrothermal conditions. It was possible for synthesis of phosphodiester bonds to occur in hot aqueous solutions (Kawamura, 2004).

Given the relevant background, it is within this context that the following thesis work is produced. Here, I have studied the interactions of multiple smaller RNA oligomers. A prebiotic system which contains small RNA oligomers that interact with each other through recombination with the help of different environmental conditions is studied herein. This work adds the prebiotic concepts for the study of RNA World and chemical evolution of small molecules. The system I chose for this thesis is the previously studied *Azoarcus* group I intron, which shows that this ribozyme can be fragmented into shorter oligomers, which then can assemble non-covalently into a catalytically active full-length ribozyme (Hayden and Lehman, 2006). This work focuses mainly on generalized RNA-directed recombination events rather than on templated-directed complementary strand synthesis. The commentary tag sequence, CAU triplet can base pair with the Internal Guide Sequence (IGS) of the *Azoarcus* ribozyme. Through a series of recombination reactions that take place with the help of IGS-tag interaction, these short oligomers can form into a fully

functional covalent ribozyme in a reaction similar to the reverse of the *in vivo* splicing reaction of *Azoarcus* group I intron. Therefore, this study showed how RNA recombination can result in the spontaneous construction of covalently contiguous group I ribozymes from short RNA oligomers in the absence of any pre-existing ribozyme. The products from its recombination reaction can catalyze further ribozyme construction, which suggests that this is a good example for an autocatalytic system. This system is a good model for the studies of the RNA World hypothesis.

The main goal of the current work was to study fragmentation and covalent self-assembly of the *Azoarcus* group I intron ribozyme from many multiple short RNA oligomers. I was particularly interested in identifying the possible IGS-tag pairs other than the natural IGS-tag pair that can be used to fragment and self-assemble the RNA oligomers. I also seek to understand how evaporation and rehydration can be used to facilitate the self-assembly reactions when many multiple RNA fragments are used in the reaction mixture.

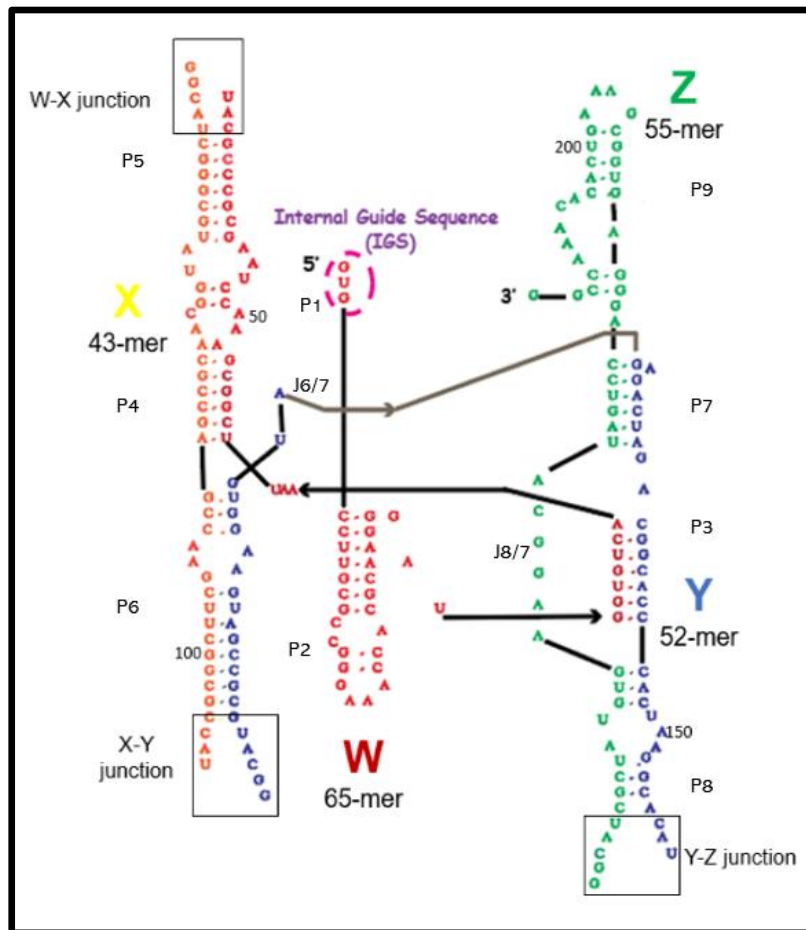
To understand the development of molecular origins on the early Earth is to understand the interactions and the thermodynamic chemistry of the small molecules. This concept is what I tried to explore in this dissertation. Chapter 2 begins with the introduction of how naturally occurring IGS-tag interaction of the *Azoarcus* group I intron can be used to fragment the entire molecule into several short fragments and help to recombine into a full-length catalytically active molecule. Chapter 3 discusses the possibility of using mutated IGS-tag interactions for fragmentation and covalent self-assembly. Chapter 4 explains how

I used different evaporation and rehydration cycles to increase the chances of self-assembly reactions through increasing concentrations of the reactants. In Chapter 5, as an augmentation to the experimental work that I discuss in previous chapters, I introduce the idea of the importance of maintaining an RNase and contamination-free laboratory when conducting experimental work with RNA and other contamination sensitive materials.

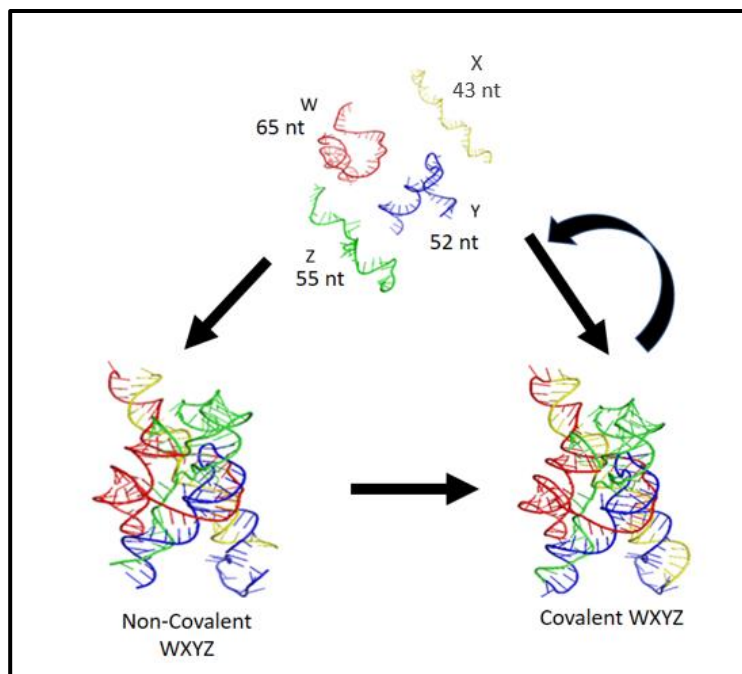
Figure 1: Breaking strategy of the *Azoarcus* group I intron.

(A) *Azoarcus* ribozyme is broken into four fragments (Hayden and Lehman, 2006): **W** (63 nt), **X** (40 nt), **Y** (52 nt), and **Z** (55 nt). All three junctions (W-X, X-Y, and Y-Z) contain head sequence (h) GGCAU on 5' end and the CAU complementary tag sequence (outlined in black) on the 3' end of the each fragment. (B) Reaction scheme of the *Azoarcus* recombinase system. First, **W**, **X**, **Y**, and **Z** fragments associate through Watson-Crick base pair interactions to form non-covalent **WXYZ** molecule. Then in the second step non-covalent **WXYZ** is catalytically connected through three recombination reactions at the **W-X**, **X-Y**, and **Y-Z** junctions.

A



B



Chapter 2

Self-Assembly of the *Azoarcus* Group I Intron from Five Small RNA Fragments

Introduction

Natural RNAs are dynamic molecules that undergo large structural changes in carrying out specific functions. Self-assembly is the autonomous organization of components into patterns or structures. This process involves different size of molecular components and various kinds of chemical interactions. Biological systems rely almost exclusively on molecular self-assembly to create complex molecular structures that carry out diverse functions. The application of self-assembly principles come from biological systems which provides ways to achieve greater complexity of organisms. Similar processes may also have played a role in the origin of life on Earth, because the earliest replication systems may have proliferated by spreading on mineral surfaces (Ferris et al.,1996). The existence of catalytic RNA suggests there may have been a mechanism through which these RNAs self-assembled from small fragments via autocatalysis.

Simple replicating molecules are likely to have played a critical role in the origin of life. Chemical systems that involve self-replication of oligonucleotides have been studied and developed as an approach to understanding the chemical processes of the origin of life on the Earth. Self-replicating molecules can be produced when two complementary subunits are covalently linked to produce a stable structure (Zielinski and Orgel,1987). Some

studies have identified the minimum requirements for molecular replication systems and showed that non-enzymatic replication is possible in a wide range of synthetic chemical systems (Luther et al., 1998). These studies provide a better understanding of the scope and limitations of self-organization process that was relevant to the origin of life on the Earth. The development of non-enzymatic self-replicating systems that are based on autocatalytic templated-directed reactions has been studied previously (Sievers and von Kiedrowski, 1994). In that study, a self-complementary template molecule AB is synthesized autocatalytically from two complementary template fragments, A and B, where these results provide a stage for the development of schemes that might explain how replicating systems based on nucleic acids arose on the prebiotic Earth.

An iterative stepwise procedure for chemical replication that exponentially increases the concentration of the oligonucleotide analogues has been studied (Luther et al., 1998). Their system employs the surface of a solid support where the precursor fragments ligate to produce the products on immobilized templates and then the ligated products are released to become new templates. The solid surface helps to separate the complementary templates which forms stable duplexes in solution. This is important because many theories of origins of biological organization focus on polymers within a range from 30-60 monomers. Ferris et al., (1996) described the abiotic synthesis of long prebiotic oligomers from activated monomers on catalytic montmorillonite surface. This provides an insight into similar types of reactions that could have taken place in the development of the RNA world.

However, these self-replicating systems have a common problem: product inhibition. These systems can suffer from high levels of product inhibition, preventing efficient replication. Some systems even do not show both self-assembly and autocatalytic properties simultaneously. Templated-directed polymerization is mainly based on the use of nucleotide triphosphates which are high energy building blocks in phosphor transesterification reactions. Fortunately, RNA recombination (as opposed to the templated-directed ligation) can address these issues. Mainly, recombination uses the exchange of nucleotides with the conversion of phosphoester bonds. For example, Group I introns catalyze the self-splicing reactions through a nucleophilic attack of the 2'- or 3'-OH group from the ribonucleotide onto a phosphodiester linkage that connects two adjacent nucleotides in the substrate. Therefore, this mechanism does not require an activated high energy bond (Lehman, 2003).

I have therefore studied the possibility of self-assembly of multiple small RNA fragments into an autocatalytic system through recombination. The system that is described herein shows the possibility of how recombination has played a direct role in self-assembling of short inactive RNA oligomers into a covalent system. Thus, this model is a great way to demonstrate how recombination played a major role in the origins of life as one of the main mechanisms by which complex phenotypes could have arisen from a collection of much smaller RNA fragments.

In one of the studies using the *Tetrahymena* group I intron, it was shown that the ribozyme can splice together multiple oligonucleotides aligned on a template strand to produce a fully complementary product strand. This reaction shows the feasibility of RNA catalyzed RNA replication (Doudna and Szostak, 1989). It was shown that the two main tertiary structure domains of the catalytic core of the *Tetrahymena* group I intron were synthesized separately and could self-assemble into a catalytically active complex. This suggests that the domains interact through multiple tertiary interactions. The organization of RNA into domains of tertiary structure can be seen in large structured molecules such as self-splicing introns, ribonuclease P, and ribosomal RNAs. This shows that a Group I intron active site can be assembled from smaller structured units of RNAs, thus raising the possibility that there may have been a primitive spliceosome and a primitive ribosome without any proteins (Doudna and Cech., 1995). Derivatives of the *sunY* self-splicing ribozymes were separated into three short RNA fragments, and they are able to form the active catalytic complexes (Doudna et al., 1991). One of these derivatives catalyzes the synthesis of the complementary RNA strand to one of its own subunits. Also, these derivatives can efficiently catalyze the synthesis of complementary strand RNA by templated-directed assembly of oligonucleotides. All these previous studies demonstrate that prebiotically synthesized oligonucleotides might have been able to assemble into complex structures which can go through self-replication. Although the these described systems provide a great insight about prebiotically how the organization of RNA into domains of tertiary structure a common theme in large RNA molecules could be, these ribozyme complements

possessed no catalytic activity, and consequently autocatalytic self-assembly was not possible.

Group I introns are removed from precursor RNAs through two consecutive transesterification reactions (**Figure 2**). Many group I introns do not require protein factors to self-splice *in vitro*, whereas others need protein factors to facilitate correct folding of the ribozyme core. A prerequisite for splicing is the binding of an exogenous guanosine (exoG) cofactor to a pocket in the catalytic core of the intron, called the G-binding site. During the first step of splicing, the cofactor attacks the 5'-splice site and attaches to the intron resulting in the release of the upstream exon. The exoG leaves the G-binding site and is replaced by the last nucleotide of the intron, which is always a G (called ω G). The second step is initiated by an attack by the 3'OH group of the released exon on the 3'-splice site, which results in ligation of the exons and release of the intron RNA. Successful catalysis is dependent on the correct folding of the intron. (Cech, 1990; Haugen et al., 2005).

The *Azoarcus* intron was crystallized in a complex representing the conformation before second step of splicing (Adams et al., 2004). Omega G (ω G) is sandwiched between adjacent purines in the major groove of pairing region 7 (P7). The 'purine sandwich' stacks smoothly with joiner regions J8/7 below and helix P9.0 above. A detailed analysis of the structural model of the *Azoarcus* intron illustrates several important features of RNA tertiary structure. Base stacking interactions coaxially align helices, shape helix interfaces

and create the guanosine-binding pocket. Nucleotides in J2/3, J3/4, J6/7 and the 5' side of P3 form a continuous belt that wraps almost entirely around the outside of the intron core. The active site and the guanosine substrate binding pocket are more complex. The extensive coordination of active site metals at the domain interface suggests that their structural and catalytic roles are virtually inseparable.

The *Azoarcus* intron is of the same group IC subclass as the more extensively investigated *Tetrahymena* intron, and both introns appear to follow an equivalent reaction mechanism. Therefore, Lehman lab studied the *Azoarcus* group I intron as the model system to addresses the autocatalytic issue in self-assembled RNA systems focusing on generalized RNA-directed recombination events, rather than on template-directed complementary strand syntheses. In the Lehman lab, the *Azoarcus* ribozyme was shown to catalyze RNA oligonucleotide recombination in a single reaction vessel (Riley and Lehman, 2003). This system demonstrates a mechanism by which covalently contiguous group I ribozymes can self-assemble from inactive RNA fragments and then catalyze further assembly reactions (Hayden and Lehman, 2006). Key characteristic of the self-assembled system that they show here is, autocatalysis. It was realized that RNA-directed recombination of inactivated oligonucleotide fragments of 40-60 nucleotides results in covalent self-assembly of these small fragments into a full- length contiguous ribozyme.

As shown in the previous studies of *Azoarcus* self-assembly (Hayden and Lehman, 2006; Draper et al., 2008), the molecule can be fragmented into four short inactive oligomers; **W**,

X, **Y**, and **Z**. These four fragments can self-assemble into a fully functional catalytically active covalent full-length molecule. The work described in this chapter shows how fragmentation into even more shorter fragments can form a catalytically active full-length ribozyme. To achieve this, I tested new break points in the fragments **X**, **Y**, and **Z** (**Figure 3**). All the attempts that were taken previously to find a new breaking location in **W** fragment were unsuccessful. Therefore, in this work, I did not test any new breaking junctions in the **W** fragment with natural IGS-tag combination, which is GUG-CAU.

In the current study along with the five-piece self-assembly reactions, I focused on two other main concepts of these recombination reactions. One is to study and optimize how to make a stable and active trans-complex in the reaction mixture. The main reason for this is because having one more extra fragment in the system cause the reactants to go through an additional chemical step. In order to overcome this issue, several reaction conditions for these self-assembly reactions were optimized, which will be discussed in chapter 4. And the other major experimental analysis was to identify which recombination mechanism took place at each junction (**Figure 4**).

Figure 2: The two-step process of group I intron self-splicing *in vivo*.

The first step is the nucleophilic attack by the 3'-OH of an exogenous guanosine cofactor, resulting in the covalent attachment of the guanosine to the 5' end of the intron and a free 3'-OH of the 5'-exon. The second step is the nucleophilic attack by this newly exposed 3'-OH, resulting in joined exons and the released intron. The result of these reactions is nucleophilic attack on a phosphodiester linkage between two nucleotides.

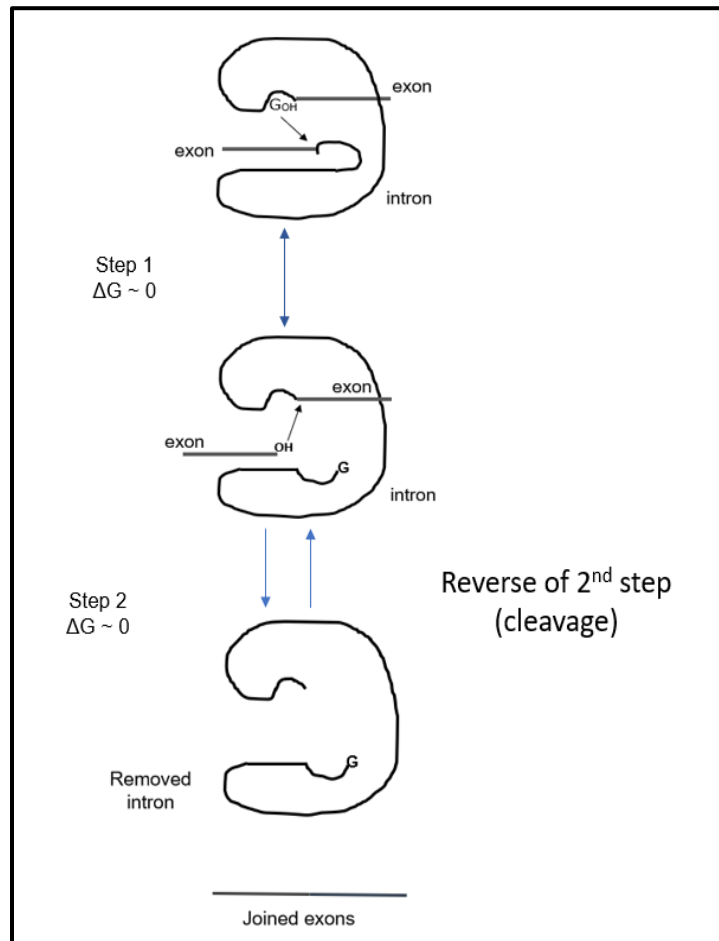
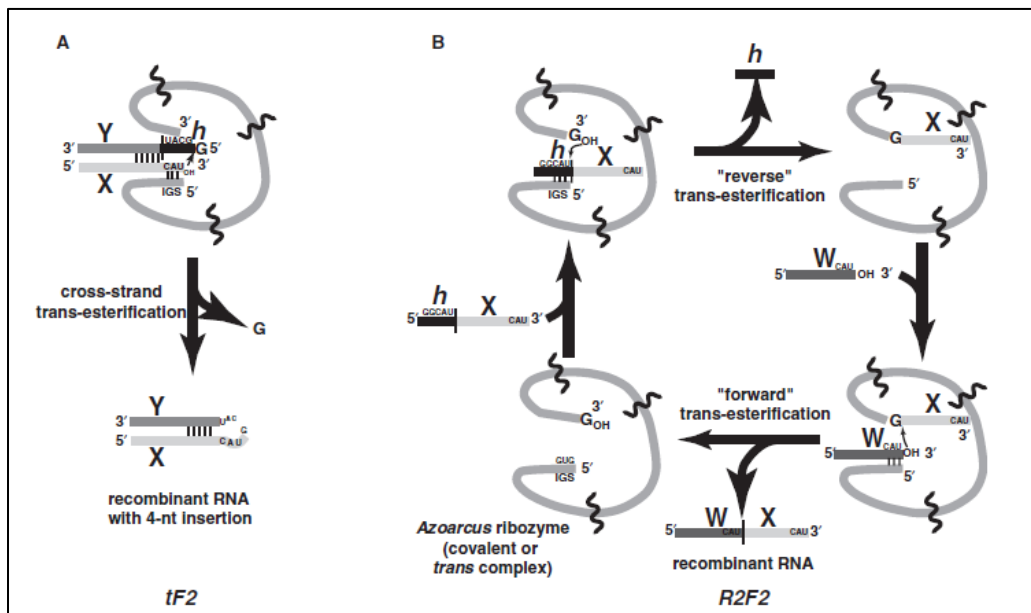


Figure 4: The two possible mechanisms of RNA-catalyzed recombination. The reaction scheme showing the two different recombination mechanisms that RNA fragments go through when the junctions are forming. (Adapted from Draper et al., 2008).

(A) The ‘*tF2*’ mechanism in which the ribozyme which can either be a covalently contiguous *Azoarcus* RNA or a trans complex comprised of 2–4 hydrogen-bonded RNA fragments (delineated by wavy lines), binds a duplex RNA and catalyzes a cross-strand attack resulting in a short insertion that is usually GCAU. Slippage of the duplex in the active site can lead to shorter (e.g. CAU) or longer (e.g. GGCAU) insertions. This reaction occurs in the forward direction during the second step of tRNA self-splicing (hence *tF2*).

(B) The ‘*R2F2*’ mechanism in which the ribozyme catalyzes two successive phosphotransfer reactions on exogenous RNA substrates resulting in a recombinant RNA molecule. Correct splicing events are guided by a specific 3 base pair interaction between the IGS (GUG) in the catalyst and its complement (CAU) in the substrates. This reaction is effectively the reverse of the second step of in vivo group I intron splicing followed by the forward of the second step (hence *R2F2*).



Results

I. Finding a breaking location in X fragment

Guided by structural data available for *Azoarcus* group I intron, and by the operative breakpoints described earlier, first I decided to analyze the five-piece self-assembly reaction by choosing a new breaking location in the **X** fragment. In the full-length **WXYZ** *Azoarcus* sequence **X** fragment is the shortest fragment of all four fragments, 43 nucleotide long. When I analyzed the nucleotide sequence of the **X** fragment, I was able to identify two main locations. The first location is C85A86A87 and second break point is located at C96A97A98. The natural complementary tag sequence for *Azoarcus* group I intron is CAU. Therefore, I decided to mutate the third adenosine in both above triplets into uracil.

When I tested the C85-A87 location, I first started with two-piece self-assembly reaction. I incubated the **WX1** and **WX2YZ** RNA fragments 1:1 molar ratio. Self-assembly reaction was carried out in EPPS pH 7.5 buffer at 48°C for 8 hours with 2μM of final RNA concentration and 100 mM final Mg²⁺ concentration. **Figure 5** shows the denaturing gel demonstrating the two-piece self-assembly described here. I observed self-assembly of covalently contiguous RNA approximately 200 nucleotides in length. The full-length product yield of this reaction was 30% after 2 hours. I then tested the three-piece self-assembly reaction with RNA fragments **W**, **X1**, and **X2YZ**. The reaction was carried out under the same reaction conditions as described above. **Figure 6** denaturing gel shows the formation of the full-length product starting from two-hour time point through eight-hour time point and I was able to observe a 23% of a product yield in this reaction after the first

two hours. With that result, I then tested the four-piece self-assembly with RNA fragments, **W**, **X1**, **X2**, and **YZ**. **Figure 7** shows the denaturing gel that resulted from this four-piece self-assembly. As shown in this figure, I observed the production of the full-length covalent **WXYZ** molecule starting from the two-hour time point. This reaction had 40% product yield two hours after the reaction has started. When I tested the five-piece self-assembly reaction with **W**, **X1**, **X2**, **Y**, and **Z** RNA fragments I was not able to detect any full-length **WXYZ** RNA produced even after 8 hours.

As the next step, I tested the second location that is located on the **X** fragment; C96-A98. First, the two-piece self-assembly reaction was tested with RNA fragments **WX2** and **X2YZ**. But even after long-term incubation, this reaction was not successful. Therefore, I did not proceed further with this location.

Figure 5: Two-piece self-assembly with X fragment: WX1+X2YZ: The two -piece self-assembly was performed by incubating 1 μ M of each fragment in 100mM MgCl₂ at 48⁰C. The size controls for **WXYZ**, **X1YZ**, and **WX2** and the time points are denoted on the gel.

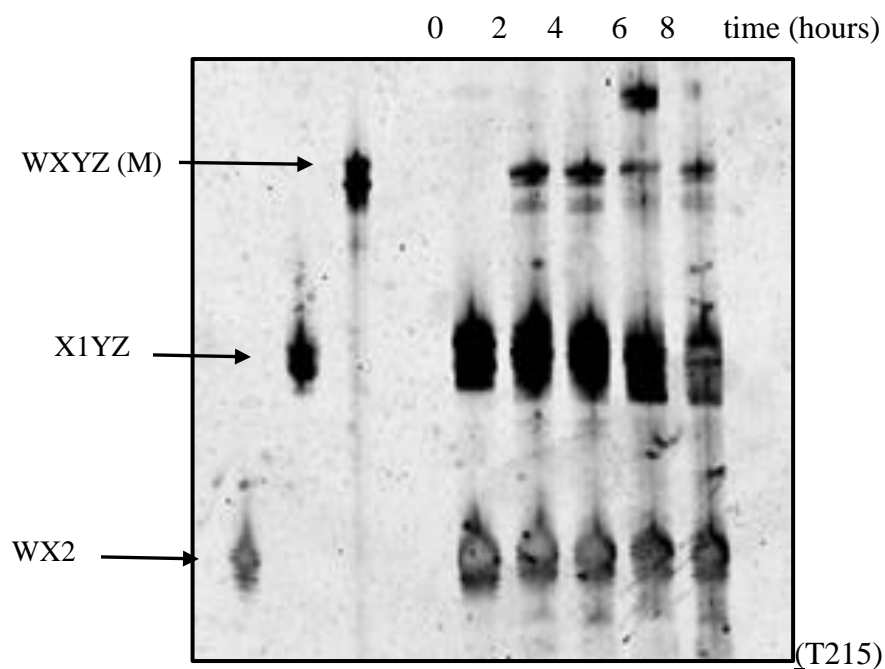


Figure 6: Three-piece self-assembly with X fragment: W+X1+X2YZ. The three-piece self-assembly was performed by incubating 2 μ M of each RNA fragment in 100mM MgCl₂ at 48⁰C. The size controls for **WXYZ**, **X₂YZ**, **W** (65 nt) and the time points are denoted on the gel. A full-length **WXYZ** RNA product has started to form from two hours of incubation.

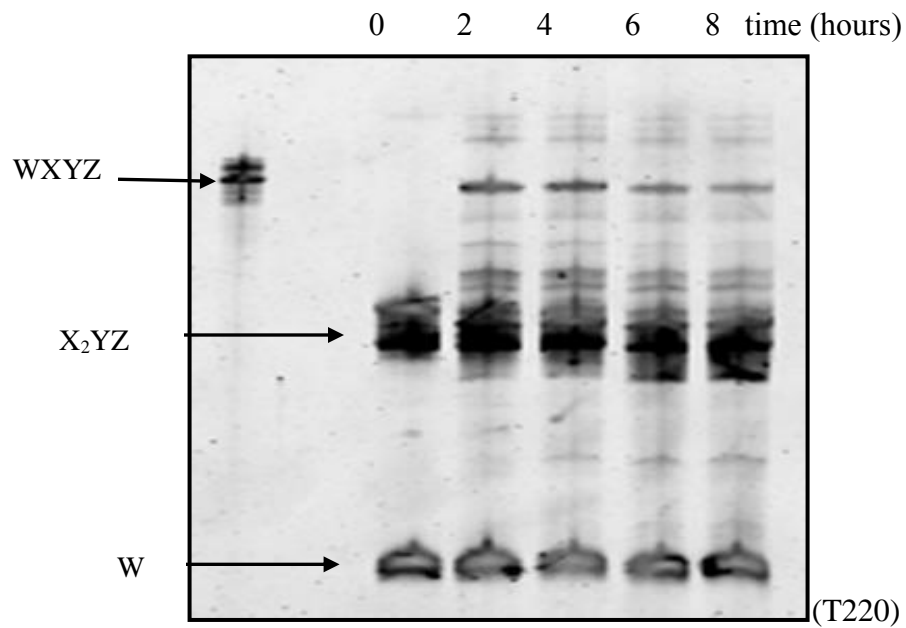
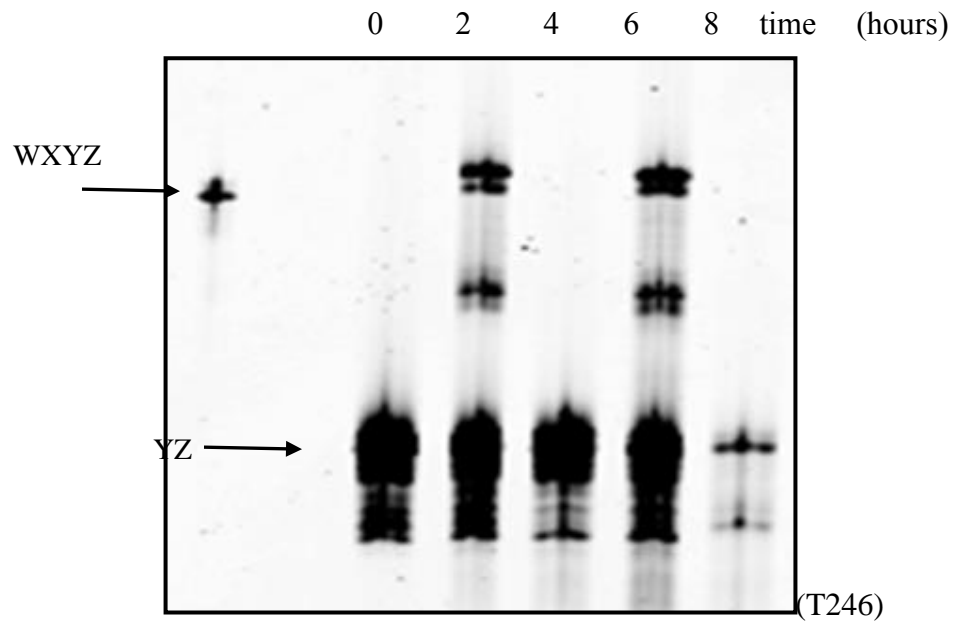


Figure 7: Four-piece self-assembly with X fragment: W+X1+X2+YZ. The four-piece self-assembly was performed by incubating 2 μ M of each fragment in 100mM MgCl₂ at 48⁰C. The size controls for **WXYZ**, **YZ** (110 nt) and the time points are denoted on the gel. As shown on the gel, a full-length **WXYZ** RNA product is present at two-hour and six-hour time points.



II. Finding a breaking location in Y fragment

The next approach was to find another feasible breaking location to obtain five-piece self-assembly. Because I had already tested the **X** fragment, thus I analyzed the **Y** fragment and selected a new breaking location within the fragment. The **Y** fragment is 52 nucleotides long and consists of the one of the main joiner regions of the *Azoarcus* secondary structure, J6/7. This joiner region contributes to two important structural elements within the intron active site (Adams et al., 2004). The new breaking location that I chose is located at U124G125U126. The nucleotides G125 and U126, form a major groove triplex with the pairing region, P4. I mutated the nucleotide U124 into C and the nucleotide G125 into A, to obtain the CAU, which is the natural tag of *Azoarcus* group I intron. This was the next new breaking location to test for the five-piece self-assembly (shown in **Figure 3**).

I first tested the two- piece self-assembly of **WXY1** and **Y2Z** to see if the new breaking location is feasible. As shown in **Figure 8**, I was able to monitor the formation of the full-length product starting from two hours and the product yield was 0.5% after four hours. Moving further, the next step was to test the three-piece self-assembly with **W**, **XY1**, and **Y2Z** RNA fragments. As shown in **Figure 9**, we saw the full-length product in three-piece assembly as well with 2.5% product yield after four hours. I was not able to see any full-length covalent **WXYZ** RNA produced in four-piece self-assembly reaction. The reason for further self-assembly reactions were not

successful at this breaking location because J6/7 sequence in the *Azoarcus* intron sequence is absolutely conserved and by mutating two of these conserved nucleotides it could be not feasible for the *trans* complex to form through initial recombination reactions (Cannone et al., 2002).

Figure 8: Two-piece self-assembly with Y fragment: WXY1+Y2Z. The two-piece self-assembly was performed by incubating 2 μ M of each fragment in 100mM MgCl₂ at 48^oC. The size control for WXYZ, WXY₁ (124 nt), Y₂Z (86 nt) and the time points are denoted on the gel.

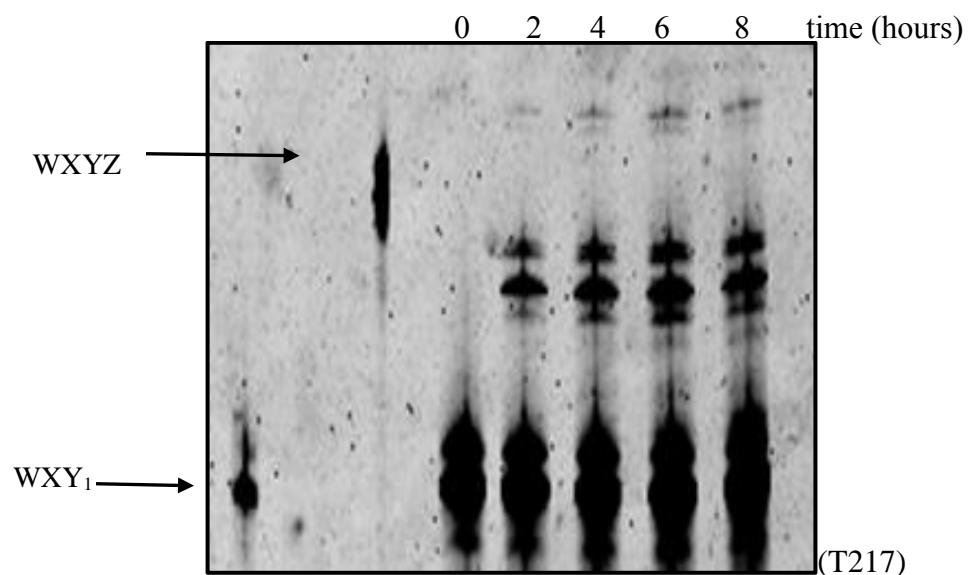
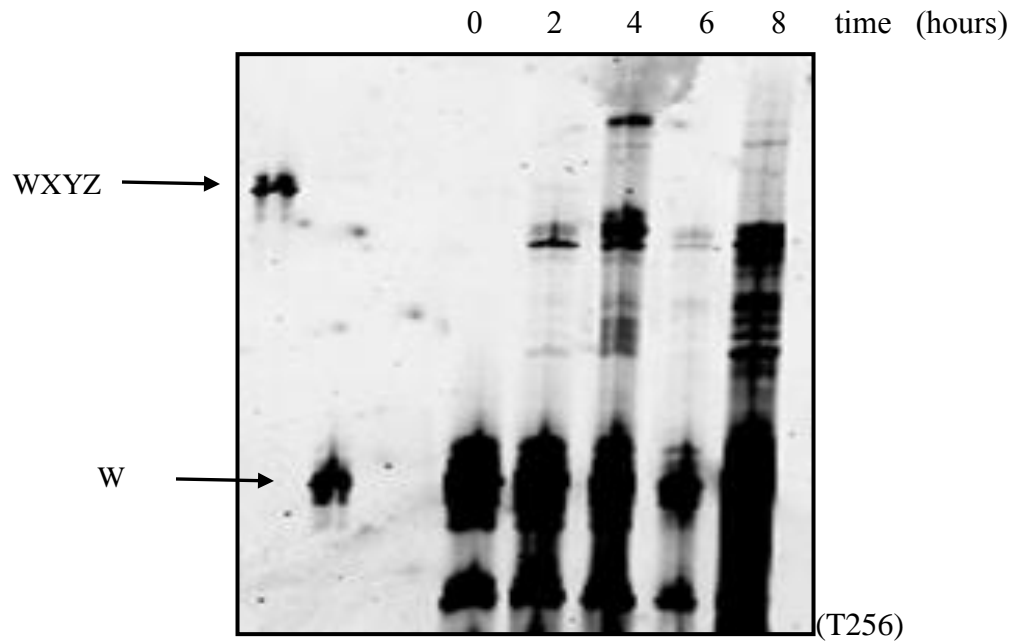


Figure 9: Three-piece self-assembly with Y fragment: W+XY1+Y2Z. The three-piece self-assembly was performed by incubating 2 μ M of each fragment in 100mM MgCl₂ at 48⁰C. The size controls for **W**, **WXYZ**, and the time points are denoted on the gel.



III. Finding a breaking location in Z fragment

When both **X** and **Y** fragments were tested it was necessary to mutate the nucleotide sequences at the new breaking locations to obtain the CAU natural complementary tag sequence of *Azoarcus* group I intron. Although binary and tertiary self-assembly reactions were successful as shown in the above sections I was not able to achieve the five-piece self-assembly reaction in both instances. Therefore, I decided to test a new breaking location where I did not have to mutate the nucleotide sequence. Because making mutations can affect the structural stability of the intron when it is being recombined with small fragments into a full-length ribozyme. Therefore, finding a location that has the naturally occurring tag sequence is important. Therefore, I consider the **Z** fragment in which it has a CAU sequence already exists in the joiner region 8/7 (J8/7) at C171A172U173 in the intron, based on the numbering system of Adams et al., (2004). (**Figure 3**). This is the only naturally occurring IGS tag sequence that the *Azoarcus* group I intron has in its structure. After identifying this location, the first step was to test whether this location is functional by performing a two-piece self-assembly with **WXYZ₁** and **Z₂** RNA fragments. In **Figure 10A**, formation of the full-length ribozyme can be seen after two hours during the reaction. This was a promising result that showed that the new breaking location in the **Z** fragment is functional and obtained a product yield of 2.5% after 4 hours. This two-piece self-assembly reaction was carried out the reaction at both 48⁰C and 60⁰C with 100mM final Mg²⁺ concentration, 2μM final RNA concentration at pH 7.5 in EPPS buffer. As shown in **Figure 10B** full-length ribozyme formed starting from two hours of incubation time at 60⁰C as well. This led to the conclusion that the new breaking junction in the **Z** fragment

is functional not only at 48°C, but also at 60°C. The next steps were to demonstrate the functionality of the new breaking location with three-piece (**W+XY+Z₁+Z₂**), four-piece (**W+X+Y+Z₁+Z₂**), and five-piece (**W+X+Y+Z₁+Z₂**) self-assembly reactions. As shown in **Figure 11, 12** and **13**, full-length ribozyme indeed has formed in these reactions. In **Figure 13**, the first five-piece self-assembly ever recorded was observed with 16% yield of the **WXYZ** product ribozyme after 9 hours is surprisingly high, given that this reaction is involved with additional fragments and the reaction pathway is slow. This result has recently been published (Jayathilaka and Lehman, 2018)

Figure 10: Two-piece self-assembly with Z fragment: WXYZ₁+Z₂. The binary self-assembly was performed by incubating 2 μ M each of WXYZ₁ and Z₂ RNA fragments in 100mM MgCl₂ at 48⁰C (A) and 60⁰C (B). The time points are shown, and the size controls are denoted on the gel for Marker (M), WXYZ, and WXYZ₁ (190 nt).

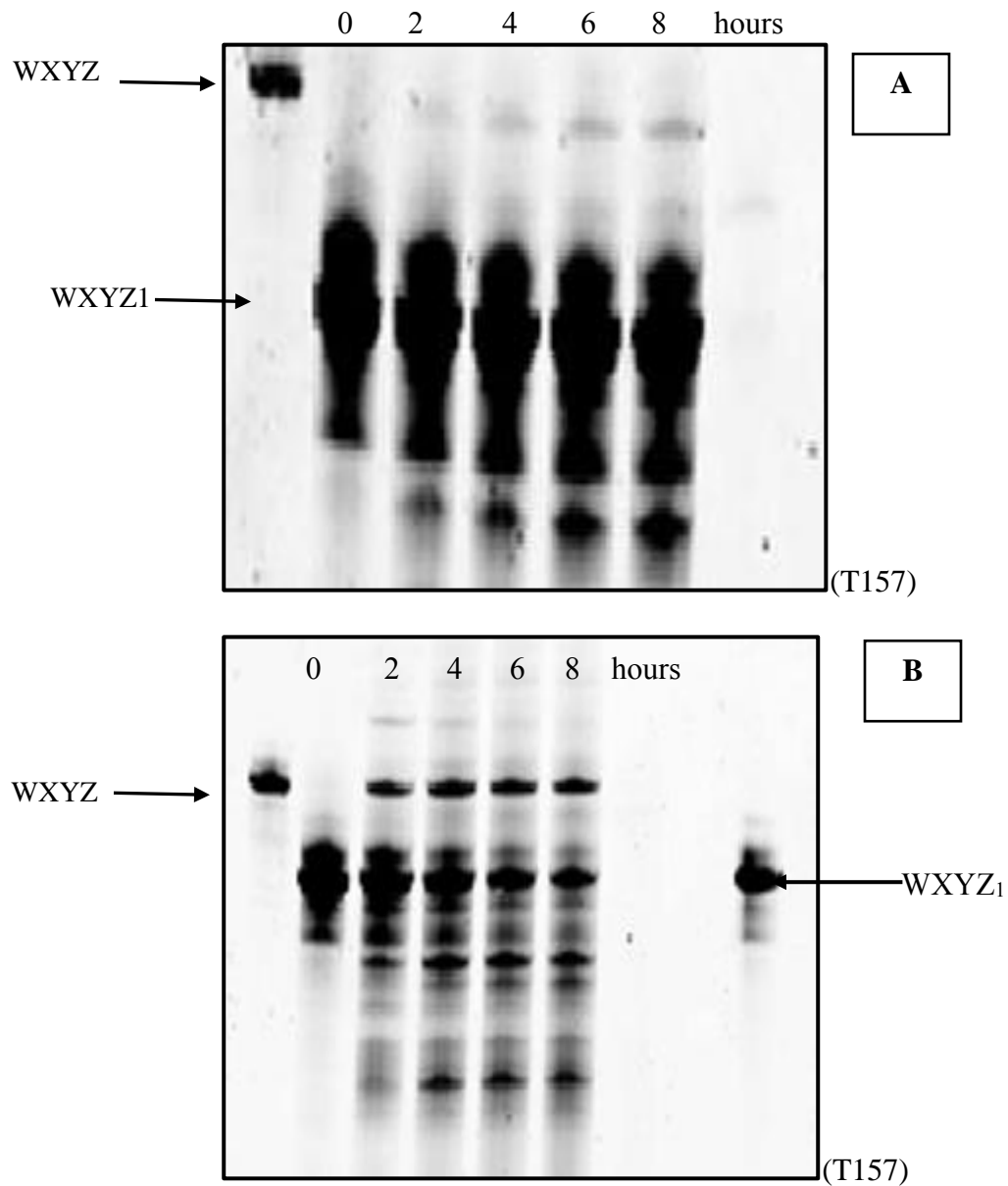


Figure 11: Three-piece self-assembly with Z fragment: W+XYZ1+Z2. The three-piece self-assembly was performed by incubating 2 μ M of each fragment in 100mM MgCl₂ at 48⁰C. The size controls for XYZ₁ (112 nt), WXYZ, and the time points are denoted on the gel.

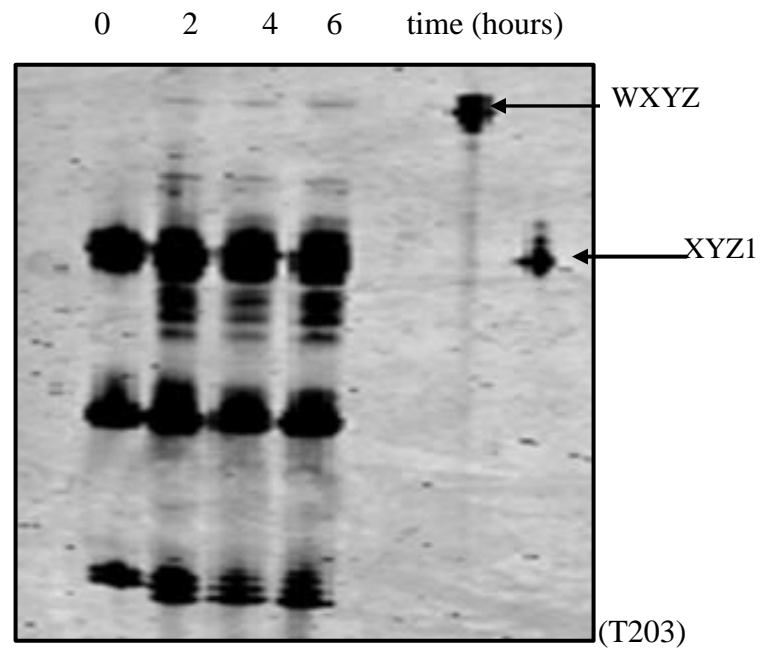


Figure 12: Four-piece self-assembly with Z fragment: W+X+YZ1+Z2. The four-piece self-assembly was performed by incubating 2 μM of each fragment in 100mM MgCl_2 at 48°C . The size controls for **XY**, **WXY** and **WXYZ** and the time points are denoted on the gel.

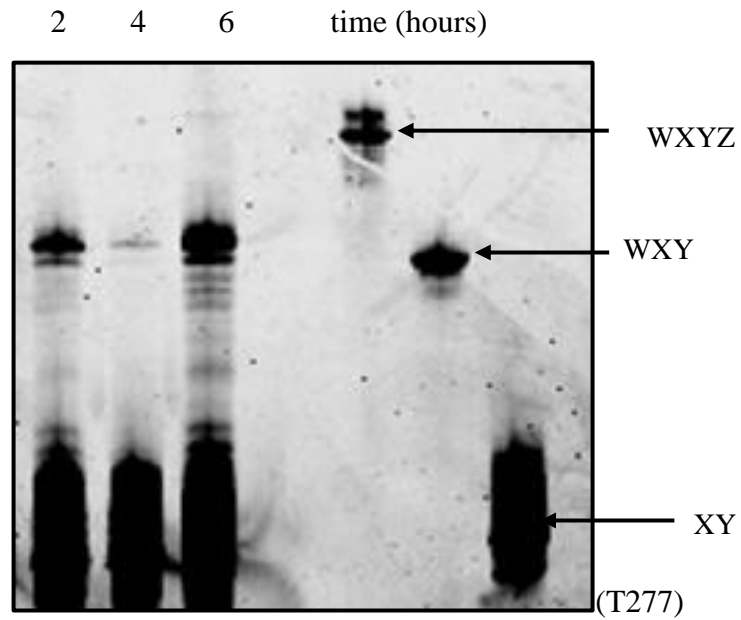
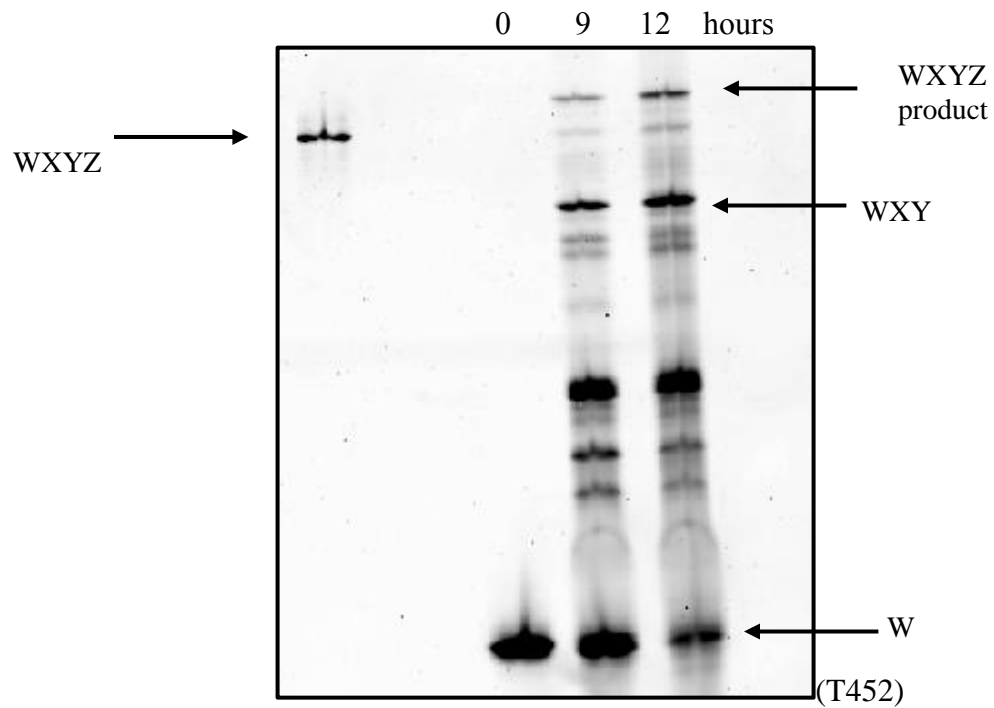


Figure 13: Five-piece self-assembly with Z fragment: W+X+Y+Z1+Z2. The five-piece self-assembly was performed by incubating 2 μ M final concentration of each RNA fragment in 100mM MgCl₂ at 48⁰C (Jayathilaka and Lehman, 2018). The size controls for **WXYZ** (220 nt) and **WXY** (157 nt) and the time points are denoted on the gel.



The next task was to determine which recombination mechanism was responsible for the recombination reaction that occurred at each of the four junctions in the five-piece assembly reaction. Previously two mechanisms had been identified (Draper et al., 2008) that could occur at a junction between two RNA fragments in these self-assembly reactions, as catalyzed by the covalent *Azoarcus* ribozyme or non-covalent *trans* assembled complex. These mechanisms were termed R2F2 and tF2. The first one is a two-step mechanism that invokes both the forward and reverse steps of the *in vivo* splicing pathway of group I introns. The second mechanism is a one-step reaction in which the ribozyme catalyzes a cross-strand attack of the 3'-G on one strand on an internal nucleotide of a hybridized strand. These two mechanisms can be distinguished, because the tF2 reaction typically inserts the 4-nt sequence GCAU at each recombination junction, with the consequence that the overall self-assembly product is slightly larger than the wild-type *Azoarcus* group I intron length. The tF2 mechanism, being a single step, is also kinetically favored, but does require some base-pairing between the two RNA fragments that are being recombined. In contrast, the R2F2 mechanism is slower, but can theoretically occur at any RNA site that contains the tag sequence (CAU). As this sequence lies at the **Z1-Z2** recombination site, and I did not specifically modify the surrounding sequences such that the tF2 mechanism would be promoted, I expected recombination at that junction to proceed through the R2F2 mechanism, but I wanted to confirm this by nucleotide sequence analysis.

I thus excised the full-length band after 12 h of the five-piece reaction, subjected it to an RT-PCR reaction specific for full-length **WXYZ**, and then cloned the DNA into an *Escherichia coli* vector and performed manual (Sanger) nucleotide sequence analysis on several individual resultant bacterial colonies. The sequences at the **W-X**, **X-Y**, **Y-Z1**, and **Z1-Z2** junctions allowed us to reconstruct which mechanisms were predominated at which junction (**Table 1**)

As expected, 100% of the reactions at the new **Z1-Z2** junction appeared to have proceeded through the R2F2 mechanism. By contrast, the **W-X** and **X-Y** junctions, which had been engineered to favor the tF2 mechanism, displayed a preference for that mechanism; although the R2F2 route can operate on any CAU-containing junction at some frequency. Moreover, should multiple reactions occur at any junction, the size of the insertion can be progressively trimmed, giving some uncertainty as to which mechanism(s) had been employed. The **Y-Z1** junction, for example, showed evidence of both mechanisms operating at high frequencies. In any event, these data confirmed our prediction that R2F2 was the dominant mechanism for the new breaking location in the **Z** fragment under investigation.

Table 1: Nucleotide sequences at the W-X, X-Y, Y-Z1, and Z1-Z2 junctions. Nucleotide sequences at the junctions in 21 cloned full-length RNAs from the five-piece self-assembly reaction. *Resulting from tF2 mechanism; others are the likely result of R2F2. [a] 76% tF2 at the W-X junction. [b] 81% tF2 at the X-Y junction. [c] 33% tF2 at the Y-Z1 junction. [d] 0% tF2 (all R2F2) at the Z1-Z2 junction. [e] These clones showed evidence of spurious recombination at the first three junctions and were counted as proceeding through the R2F2 mechanism.

Clone	W-X ^[a]	X-Y ^[b]	Y-Z1 ^[c]	Z1-Z2 ^[d]
1	GCAUCAU*	GCAUCAU*	CAU	CAU
2	GCAUCAU*	GCAUCAU*	GCAUCAU*	CAU
3	GCAUCAU*	GCAUCAU*	GCAUCAU*	CAU
4	GGCAUCAU*	GCAUCAU*	CAU	CAU
5	GCAUCAU*	GCAUCAU*	CAU	CAU
6	GCAUCAU*	GCAUCAU*	CAU	CAU
7	GCAUCAU*	GCAUCAU*	CAU	CAU
8	CAU	GCAUCAU*	GCAUCAU*	CAU
9	CAU	GCAUCAU*	CAU	CAU
10	GCAUCAU*	GCAUCAU*	CAU	CAU
11	GCAUCAU*	GCAUCAU*	CAU	CAU
12	GCAUCAU*	GCAUCAU*	GCAUCAU*	CAU
13 ^[e]	CAUCAUCAU	CAUCAUCAU	CAUGCAU*	CAU
14 ^[e]	CAUCAUCAU	CAUCAUCAU	CAUGCAU*	CAU
15	CAU	GCAUCAU*	CAU	CAU
16	GCAUCAU*	GCAUCAU*	CAU	CAU
17	GCAUCAU*	CAU	CAU	CAU
18	GCAUCAU*	GCAUCAU*	CAU	CAU
19	GCAUCAU*	GCAUCAU*	CAU	CAU
20	GCAUCAU*	GCAUCAU*	CAU	CAU
21	GCAUCAU*	GCAUCAU*	GCAUCAU*	CAU

Discussion

The *Azoarcus* group I intron genome consists of a larger number of secondary and tertiary structure elements. These structural elements are so important for the reactivity of the ribozyme. The breaking location that I chose in the **X** fragment, CAA, was mutated into CAU, which is the natural identification tag for the Internal Guide Sequence (IGS) of the *Azoarcus* group I intron. The 5' splice site of the *Azoarcus* group I intron is specified by an invariant wobble mis-pair between at the exon and a G within the IGS. The G.U wobble pair is recognized by the conserved A-rich J4/5-J5/4 symmetric loop, using the wobble receptor motif. The conserved As in the loop are mis-paired such that one adenine from each strand (J4/5 residue A58 and J5/4 residue A87) is presented on the minor groove surface to contact the wobble pair. There is an A.A mis-pair in the second helix. The breaking location that I chose in the **X** fragment was A87 located in J5/4 joiner region. Because of the importance of the base pair interactions in this adenine rich region, a single nucleotide mutation could disrupt or lower the reactivity of the ribozyme. This could be a reason for the five-piece self-assembly of the **W**, **X1**, **X2**, **Y**, and **Z** was not observed.

The length of the J6/7 (**Figure 1**) is conserved, but not the sequence. It always includes three nucleotides. The short length of J6/7 combined with the large distance that it must traverse across the active site, pulls G128 out of the P7 helix and into the P7 major groove where it forms the lid to the G-binding site (**Figure 1**). Mutagenesis in this region confirmed the functional importance of these three bases and led to

proposals for how they may interact (Michel and Westhof, 1990). With all the studies related to the structure and the sequence of the *Azoarcus* group I intron, it is confirmed that the mutation G125 to an adenine could have affected the functionality of the ribozyme, which made it difficult for the self-assembly of the multiple fragments at this new breaking point in the **Y** fragment.

The **Z** fragment is in the P3-P9 pairing region of the *Azoarcus* group I intron secondary structure. In comparison to the other elements, this region has relatively few tertiary interactions. This deficit is given its irregularly bent shape and the way it appears to wrap around the P4-P6 helix. There are two primary sites of tertiary interaction and a few isolated contacts. Isolated tertiary contacts include stacking of J8/7 residue G170 at the interface between P3 and P7 and G binding site with the G130-C177 pair in P7. C171, A172, and U173 (CAU) is in this J8/7 region (**Figure 1**). This is the only location inside the *Azoarcus* group I intron where its natural tag is located. These flanking tetraloop-tetraloop receptor contacts and the high guanosine-cytosine content are likely to be primary factors in the thermostability of this intron. Therefore, I did not have to mutate any other place within the **Z** fragment to find a new breaking point. Not having to mutate or engineer the CAU tag sequence could have been the reason for this location having been the most feasible for self-assembly of the multiple small RNA fragments.

Out of all four fragments, **W** is the longest fragment (65 nucleotides). The outermost ends of the tertiary interfaces of the *Azoarcus* group I intron are stabilized by two GAAA tetraloop-tetraloop receptor contacts. One of these is located on the L2 tetraloop and a receptor in J8. This loop region could have been an ideal location for a new breaking location, but as explained above, the importance of its tertiary interactions for the stability of the intron, I decided not to disrupt the sequence of this region to use for self-assembly reactions. In previous studies about GNRA loops it has been shown that the ability of GAAA tetraloops participate in RNA-RNA tertiary interactions, and this could be a major reason for their selective advantage (Michel and Westhof, 1990)

This system demonstrates the emergence of autocatalysis from a self-organizing anabolic reaction network involving five individually non-catalytic RNA molecules (Jayathilaka and Lehman, 2018). Moreover, this current work provides an example of predicted links between the recombination and the shorter RNA oligomer templates, which leads towards more complex RNA World. This would allow the opportunity to demonstrate the first evolving system comprised entirely of RNA, which uses autocatalysis as a preferred pathway.

Materials and Methods

RNA fragments

All the RNA fragments that are mentioned in this chapter have a CAU-containing short 5 nucleotide long sequence at their 5' end; 5' head (h=GGCAU) and at 3 nucleotide long tail with CAU at their 3' end. These two short sequences allow the RNA fragments to go through one of the recombination mechanisms that is shown in **Figure 4**.

E.g.; Z1 fragment: GGCAUCGCUAUGGUGAAGGCAU: hZ1_{CAU}

W: hW_{CAU}

X: hX_{CAU}

Y: hY_{CAU}

Z2:hZ2

RNA synthesis

RNA fragments of the *Azoarcus* ribozyme were prepared by T7 RNA polymerase run-off transcription from double-stranded DNA templates made by a PCR based synthesis techniques. All RNA fragments were gel purified using 8M Urea/ 8% Polyacrylamide denaturing gels.

Radiolabeled fragments (W containing RNA fragments) were labeled on the 5' end after transcription using OptiKinase and [$\gamma^{32}\text{P}$]•ATP.

Self-assembly reactions

RNA fragments of the *Azoarcus* ribozyme were mixed in water in amounts that would allow a final concentration of 2.0 μM . only the 5' end of the oligo corresponding to the 5' end of the *Azoarcus* ribozyme was labeled with ^{32}P . the reaction was initiated by adding the self-assembly buffer with a final concentration of either 100 mM or 25 mM MgCl_2 and 30 mM EPPS (pH 7.5) and were immediately incubated at 48°C in a heat block. At the desired times, aliquots of the reaction mixture were removed and quenched with a gel loading dye containing Urea, Xylene cyanol, Bromophenol blue, and EDTA to chelate all Mg^{2+} . Radiolabeled products were separated on an 8% denaturing polyacrylamide gel and viewed by phosphoimaging. Gels that were not radiolabeled were stained using SYBR Green II and viewed using Typhoon Trio variable mode imager.

Genotyping

WXYZ product bands from the five-piece self-assembly reaction with the **Z** fragment were excised from the 8M Urea/ 8% polyacrylamide denaturing gel and eluted by the crush and soak method. The extracted RNA was subject to RT-PCR amplification with a RT primer (5'-CCGGTTTGTGTGACTTTGCC-3') complimentary to the 5' end of the *Azoarcus* ribozyme and a second primer (5'CTGCAGAATTCTAATACGACTCACTATAG GCCTTGCGCCGGAA) for the PCR that was complimentary to the 5' end of the *Azoarcus* ribozyme plus a promoter sequence for T7 RNA polymerase. Amplification with these primers yields and ~230 bp PCR product, which was confirmed using a 50 bp genomic

ladder and visualizing through a UV lamp. Then the PCR product DNA was cloned pJET 1.2 blunt end vector and transformed into *E.coli* competent cells. Bacterial cells were plated on LB-Agar plates and grew overnight at 37°C. Then the successful colonies were subjected to colony PCR. The product DNA was then subjected to nucleotide sequence analysis using Sanger sequencing. As shown in **Table 1**, 21 resultant nucleotide sequences were compared against the wild-type nucleotide sequence of the *Azoarcus* ribozyme using BioEdit sequence alignment editor to identify which RNA-recombination mechanism was utilized at each junction during the catalysis process.

Table 2: RNA oligomers used in self-assembly reactions using GUG-CAU combination. Different RNA oligomers used in self-assembly reactions described in this chapter and their lengths in nucleotide units.

	RNA oligomer	Length (nt)
X fragments	W	65
	X1	19
	X2	24
	Y	52
	Z	55
	WX1	84
	X2YZ	131
	YZ	107
Y fragment	WXY1	129
	Y2Z	86
	XY1	64
Z fragment	WXYZ1	182
	Z1	22
	Z2	33
	XYZ1	117
	XY	95

Chapter 3

RNA Self-Assembly Experiments with Multiple Different IGS-Tag Combinations

Introduction

Group I introns fold into their secondary and tertiary structures and form an active site to catalyze their own RNA splicing *in vivo*. The significance of the conserved nucleotides within base-paired duplexes has been studied (Cech, 1988). Group I introns catalyze the phosphoester transfer reactions at the splice site, and this splice site selection relies on base-pairing interactions between the 5' end of the P1 of the intron and the complementary nucleotide region of the 3' portion of the 5' exon. As mentioned previously, these two regions are called the IGS and the “tag” respectively. Base pairing interactions between the IGS and tag mostly depend on the canonical Watson-Crick base pairing interactions. However, when the catalytic activity of the *Tetrahymena* group I intron was studied using site-specific mutagenesis, the splicing activity of the intron was lowered when a single nucleotide in either 5' exon or 3' end of the complementary tag was mutated. This shows that the substrate specificity of a group I intron can be changed by favorable Watson-Crick base pairing (Been & Cech, 1986). The 3' end of the tag always tends to form a wobble pair with the IGS. For example, the IGS of the group I intron in the tRNA^{Ile} in *Azoarcus* forms a 5'G-U3' wobble pair between the IGS and the tag sequences (Reinhold-Hurek & Shub, 1992). Kinetic analysis and pH dependence studies of the *Azoarcus* group I intron

have shown that the chemical cleavage step is the rate-limiting step, not the product release step (Kuo et al., 1999). By contrast, in the intervening sequence of the ribosomal RNA precursor present in *Tetrahymena*, product release is the rate limiting step (McSwiggen, & Cech, 1989).

Cleavage at the correct site requires the formation of the correct secondary structure. The G-U wobble base pair, which is located within the IGS-tag interaction window, can greatly influence the splicing efficiency. Mutational analysis studies showed that integrity of the P1 stem is very important for catalysis of the intron. A single nucleotide mutation in this stem could greatly reduce the splicing activity (Been et al., 1987). The bottom of the P1 stem interacts with a part of the ribozyme that is loosely fixed with relative to the active site. However, the identity of the bottom base pair does not affect reactivity greatly. This is because of the P1 stem is coaxially stacked on some stem in the enzyme core. There are at least four aspects of the RNA structure that contribute to the splice-site specificity: i) proper folding of the RNA to form the P1 stem; ii) the IGS-tag base-pair environment; iii) recognition of the G-U wobble base pair at the splice site; and, iv) the base pair above the reactive phosphate, which affects the rate of the cleavage reaction (Doudna et al., 1989)

When the sugar-phosphate backbone conformation at the wobble position in the P1 stem is altered, there are at least two possible ways to achieve the splice-site specificity. One way is to position the pyrimidine 3'-phosphate of the wobble base pair to make contact with the active site of the intron. The other possibility is that the in-line attack of the guanosine 3' hydroxyl displaces the 5' exon that is facilitated by G-U wobble base pair

(McSwiggen & Cech, 1989). A previous study showed that the *Tetrahymena* intron can insert itself into heterologous RNAs (Woodson & Cech, 1989). Reverse reaction of the self-splicing was favorable in high RNA and magnesium concentrations, and at higher temperatures. In this study they showed that the 5' exon can base-pair with the IGS of the intron, whereas the 3' exon sequences were not essential.

A previous study from the Lehman lab showed that the canonical IGS complement tag sequence for *Azoarcus*, CAU, is not strictly required for the self-assembly of the intron from its shorter fragments (Draper et al., 2008). Mutations at the 3' end of the tag sequence are well tolerated by the system, while mutations at the 5' end were deleterious. All mutations, however, lowered the overall yield of the self-assembly reaction. In the previous chapter I used the *Azoarcus* group I intron, with 5'-GUG-3' as its IGS.

Results

1. Self-assembly reactions with mutated IGS-tag combinations

I alter the system to test the hypothesis that if the canonical IGS were mutated, it can still be functional. I investigated the different self-assembly pathways by using different breaking junctions. In the interest of making several new different breaking junctions, I mutated the middle nucleotide and the 3' nucleotide of the IGS (**Table 3**). By changing the middle and the 3' end nucleotide of the natural IGS (GUG), I was able to design 15 different Internal Guide Sequences. This provided the opportunity to test the self-assembly reactions for completely new complementary tag sequences. Out of these 15 new IGS-tag combinations, only ten pairs were tested. The number six pair in **Table 3** is the natural IGS-tag pair. Self-assembly reactions that were tested and those were successful using this pair were discussed in Chapter 2. Other pairs could not be tested, either because those tag sequences were not found within the *Azoarcus* structure because they were located closer to the catalytic site of the ribozyme, or because it was difficult to synthesize the required RNA fragments through run-off transcription. These pairs are underlined in **Table 3**. This gave the opportunity to find different complementary tag triplets that are located within the *Azoarcus* intron. Then I used the 3' end of these new tag sequences as the new breaking junction. The results, presented in the next section and summarized in **Table 3** below, show that the canonical IGS complement triplet CAU of the *Azoarcus* ribozyme is not strictly required for the self-construction of this RNA from component fragments.

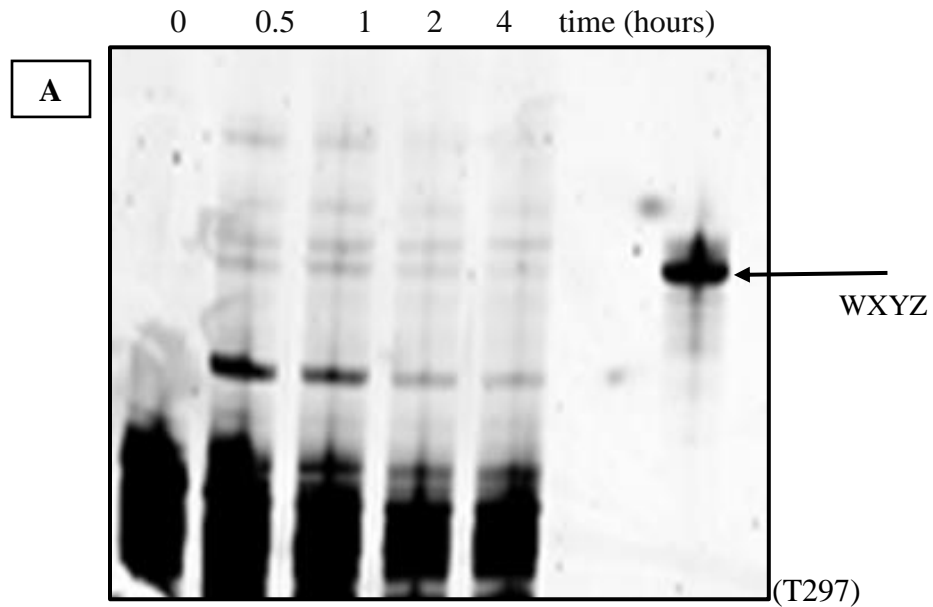
Table 3: Different approach to test self-assembly reactions with other IGS-tag combinations. Possible Different IGS-tag pairs when the middle and the 3' end nucleotide of the canonical IGS are mutated. Underlined pairs were not tested for any self-assembly reaction (see discussion). GUG-CAU is the canonical IGS-tag pair for the *Azoarcus* group I intron.

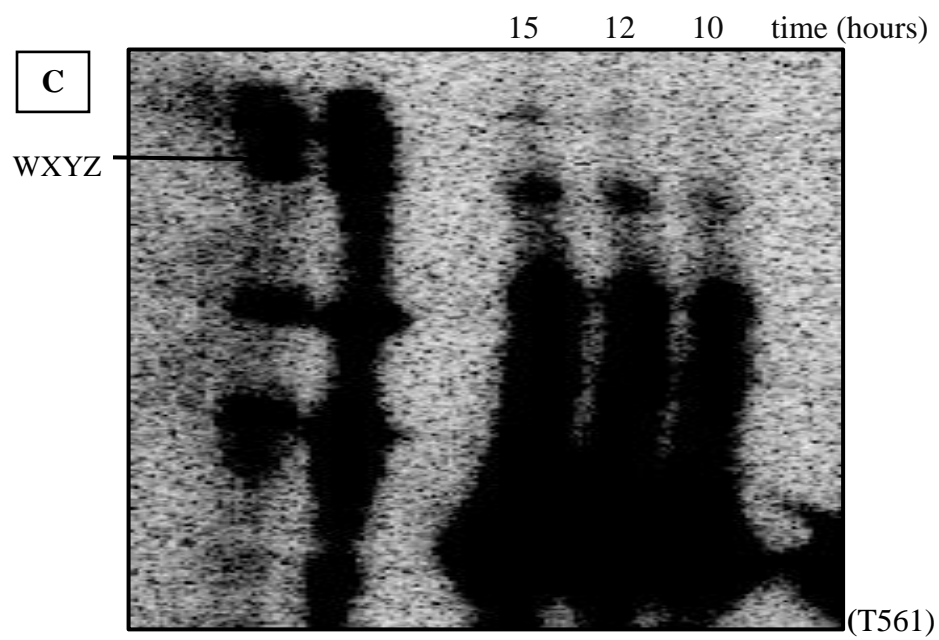
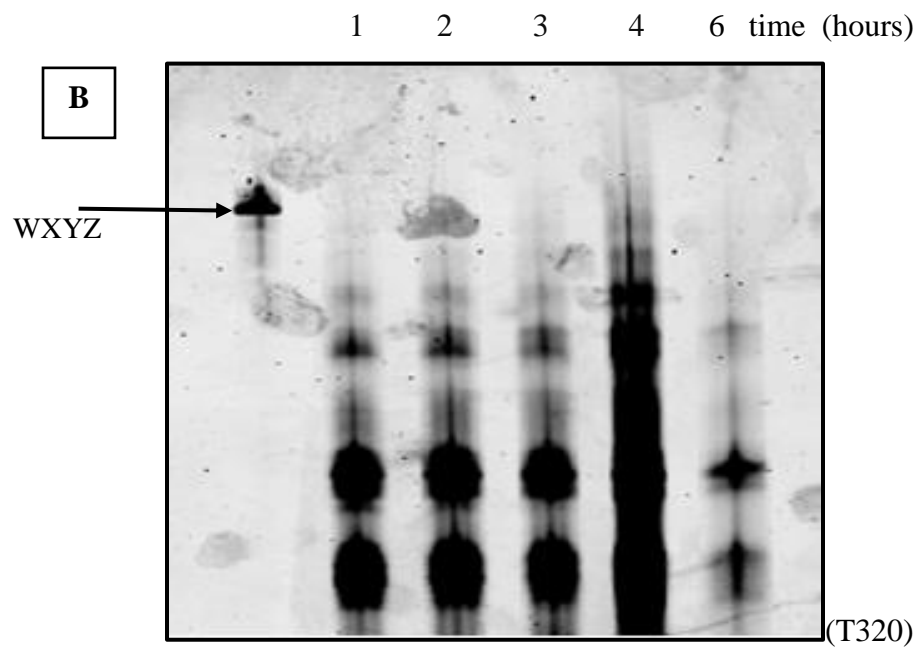
IGS (5'-3')	Complementary tag sequence (5'-3')	Highest self- assembly level observed	Maximum product yield %
<u>GCU</u>	<u>AGU</u>	<u>Not tested</u>	
GCG	CGU	5-piece	0.5%
GCC	GGU	3-piece	6%
GCA	UGU	3-piece	20%
GUU	AAU	4-piece	4%
GUG	CAU	5-piece	16%
GUC	GAU	3-piece	6%
GUA	UAU	3-piece	2%
<u>GGU</u>	<u>ACU</u>	<u>Not tested</u>	
GGG	CCU	4-piece	7%
GGC	GCU	2-piece	9%
<u>GGA</u>	<u>UCU</u>	<u>Not tested</u>	
GAU	AUU	2-piece	6%
<u>GAC</u>	<u>GUU</u>	<u>Not tested</u>	
GAG	CUU	2-piece	8%
<u>GAA</u>	<u>UUU</u>	<u>Not tested</u>	

I Self-assembly at the GCG-CGU junction

The tag sequence, CGU is located in the **X** fragment. First, the two-piece self-assembly (**WX1+X2YZ**) was tested at this junction. After the first 30 minutes the product yield was 2% (**Figure 14A**). Then the four-piece assembly (**WX1+X2+Y+Z**) was tested, and it was successful (**Figure 14B**). Lastly, I tested the five-piece assembly (**W+X1+X2+Y+Z**) and as can be seen in **Figure 14C**. This reaction was successful after incubating the RNA fragments for 15 hours. Even though, the two-piece reaction was a fast-self-assembly reaction, both the other reactions required long-term incubations.

Figure 14: Self-assembly reactions with GCG-CGU combination. **A)** Two-piece self-assembly reaction (**WX1+X2YZ**). **B)** Four-piece self-assembly (**WX1+X2+Y+Z**). **C)** Five-piece self-assembly reaction (**W+X1+X2+Y+Z**).





I. Self-assembly at the GCC-GGU junction.

The complementary tag sequence GGU can be found in three fragments: **W**, **X**, and **Z**. Two-piece self-assembly of this combination was tested at two different locations, one in the **Z** fragment (**Figure 15**) and then one in the **Y** fragment. The two-piece self-assembly was only successful in the breaking junction which is located in the **Z** fragment. After seven hours with the **Z** fragment I was able to observe a 3.5% yield of the product in this reaction. The next step, three-piece self-assembly, was carried out with the fragments **WX+YZ₁+Z₂** (**Figure 16A**) using a break within the **Z** fragment at nucleotide position U165. With the successful of 6% product yield after six hours from this three-piece self-assembly reaction, I then tested the four-piece self-assembly with the RNA fragments; **W+X+YZ₁+Z₂** (**Figure 16B**). This four-piece assembly was not completely successful to make the full length **WXYZ** *Azoarcus* ribozyme as no full-length product was detected after seven hours. Hence, I did not proceed further with this junction.

Figure 15: Two-piece self-assembly reactions with GCC-GGU combination. Fragments **WXYZ₁** and **Z₂** were incubated. Formation of the full-length product has started from three hour time point.

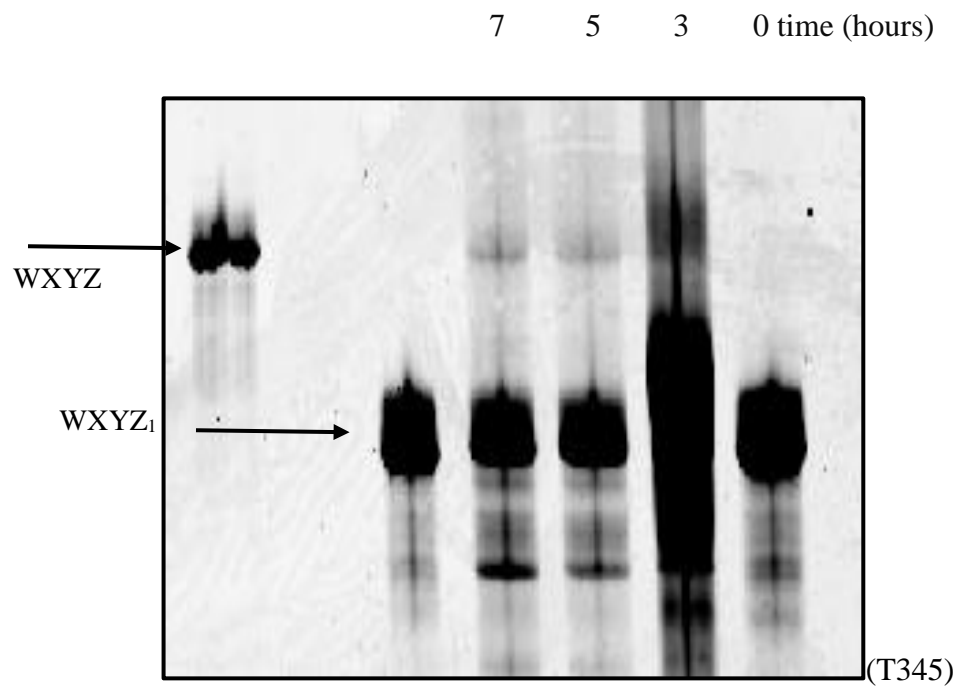
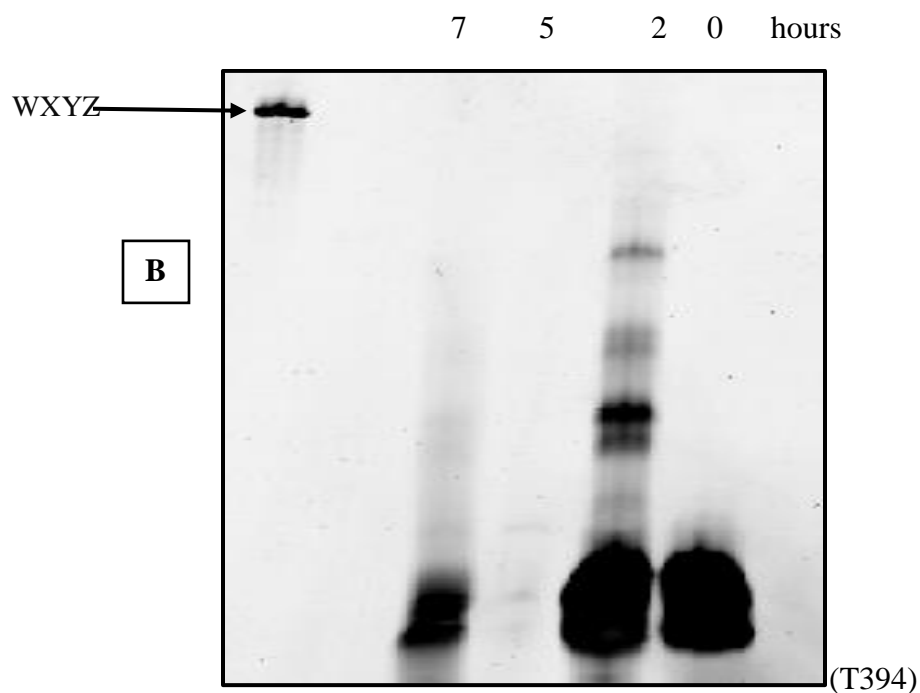
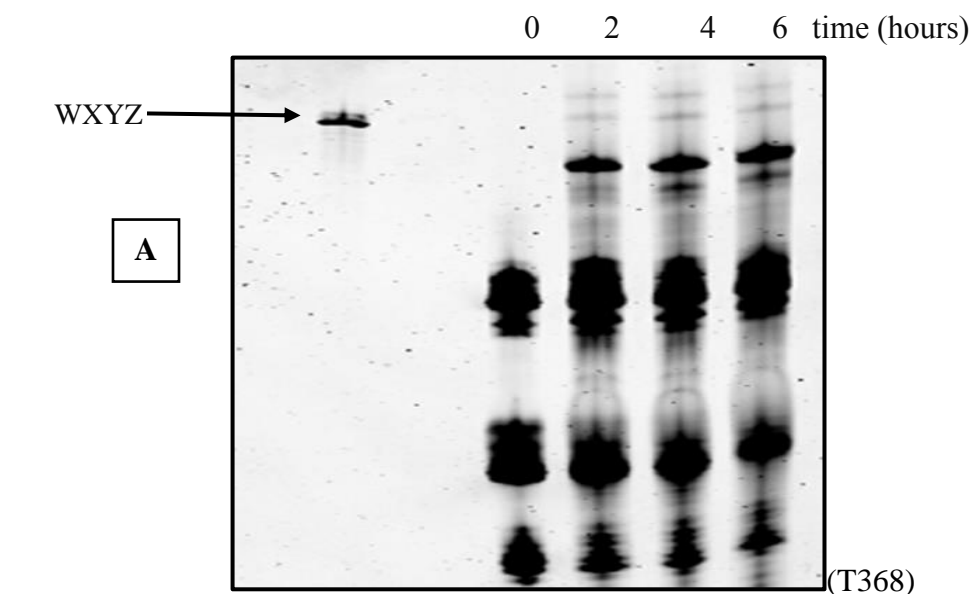


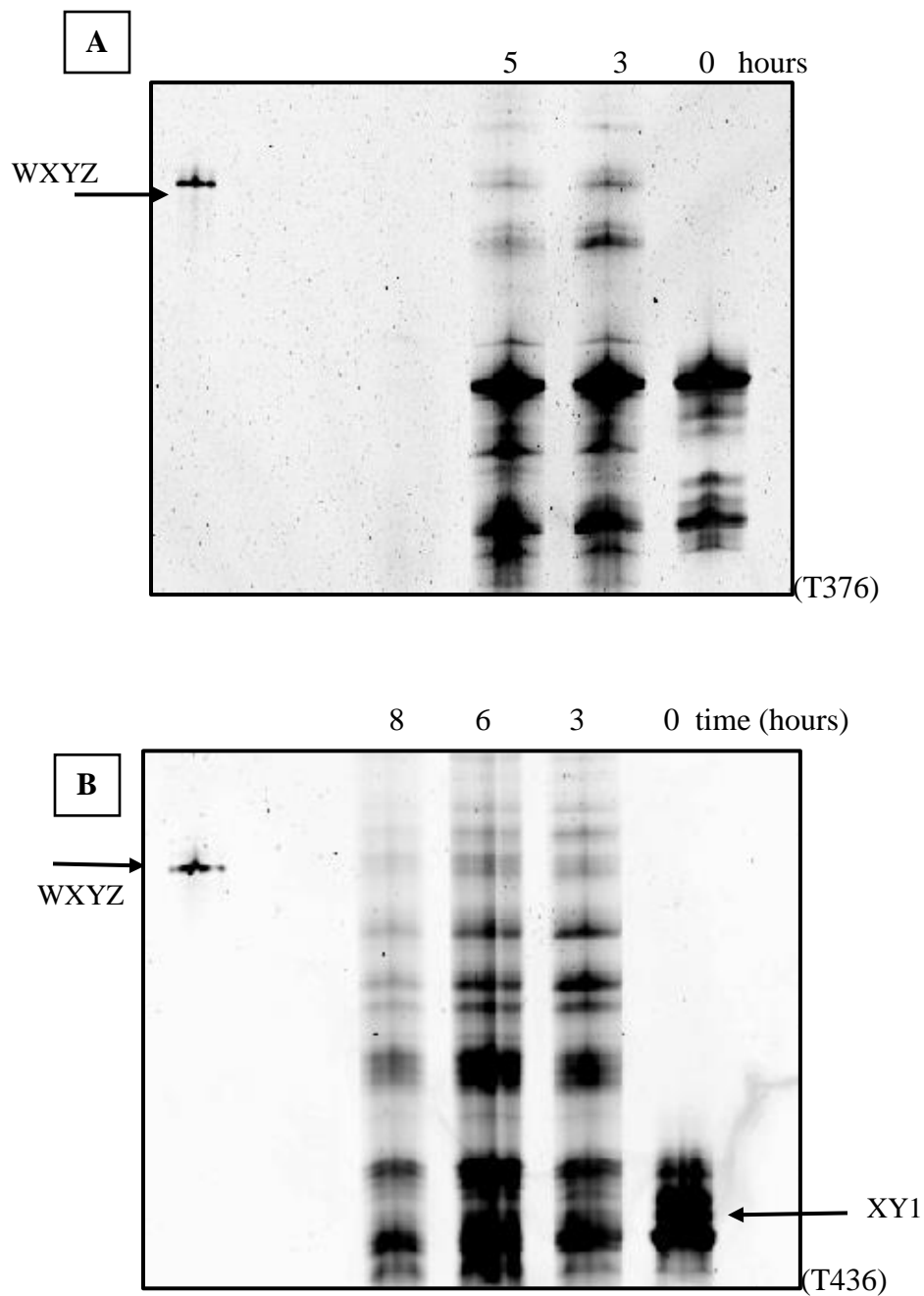
Figure 16: Three and four-piece self-assembly reactions with GCC-GGU combination. **A)** With GCC-GGU combination, after tested binary self-assembly, a three-piece self-assembly reaction (**WX+YZ1+Z2**) was carried out for 6 hours. Formation of the full-length product can be seen starting from 2 hours of incubation period. **B)** Finally, the four-piece assembly (**W+X+YZ1+Z2**) was performed.



II. Self-assembly at the GCA-UGU junction.

This new break point can be found naturally in both **W** and **Y** fragments, at positions U43G44U45 and U124G125U126 respectively. I first decided to test the self-assembly reactions with the **Y** fragment. When a two-piece self-assembly reaction was performed with the fragments, **WXY₁+Y₂Z**, I was able to observe the formation of a full length **WXYZ** product starting from 3 hours of incubation and the product yield was 30% (**Figure 17A**). With this promising result, I then proceeded to test the three-piece self-assembly with **W+XY₁+Y₂Z** fragments (**Figure 17B**) and this reaction had a 20% product yield after six hours.

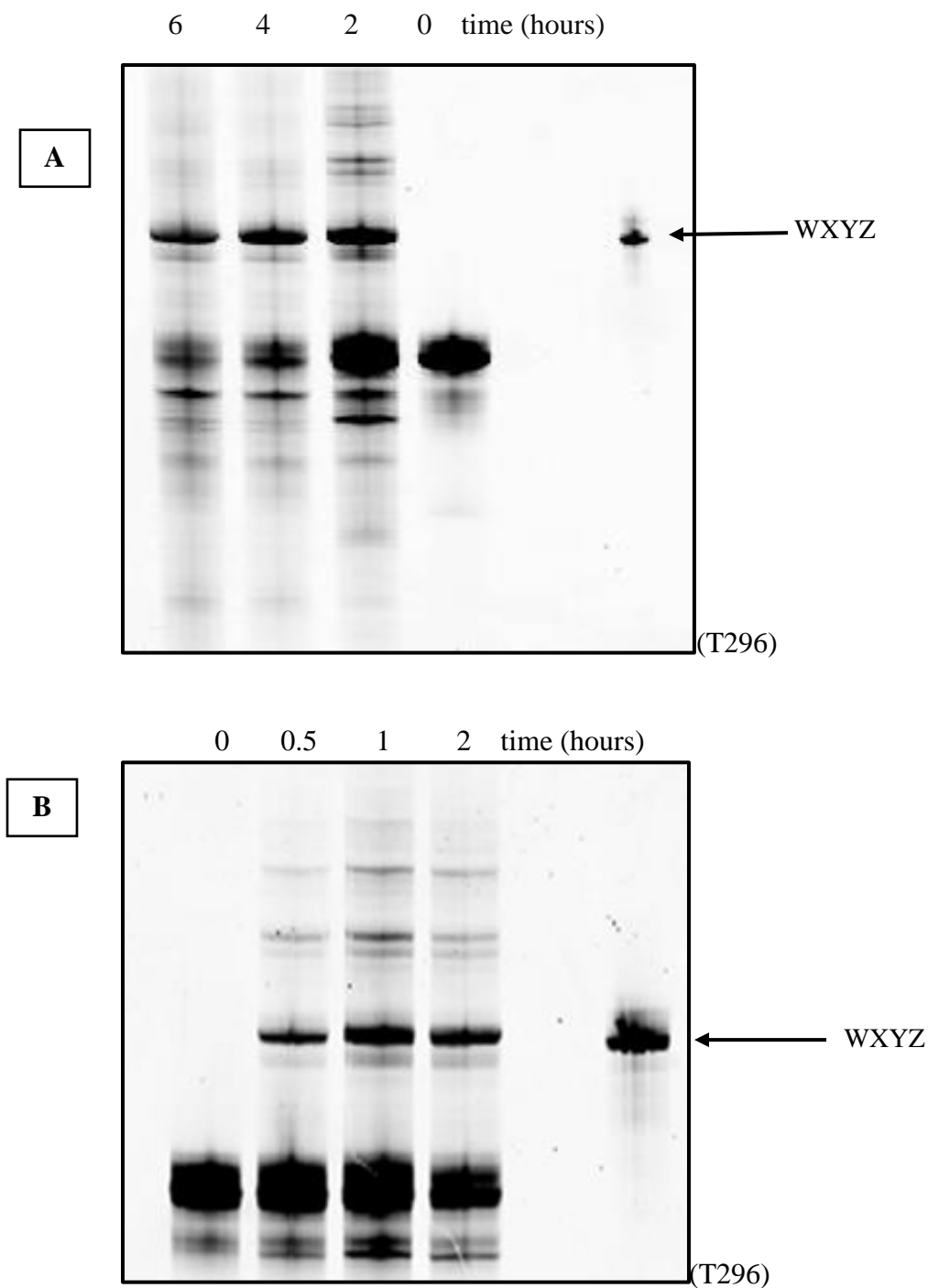
Figure 17: Self-assembly reactions with GCA-UGU combination. Molecular self-assembly of multiple RNA fragments using the GCA-UGU IGS-tag pair. **A)** Two-piece assembly using WXY_1 and Y_2Z . **B)** Three-piece self-assembly reaction among W , XY_1 , and Y_2Z fragments.

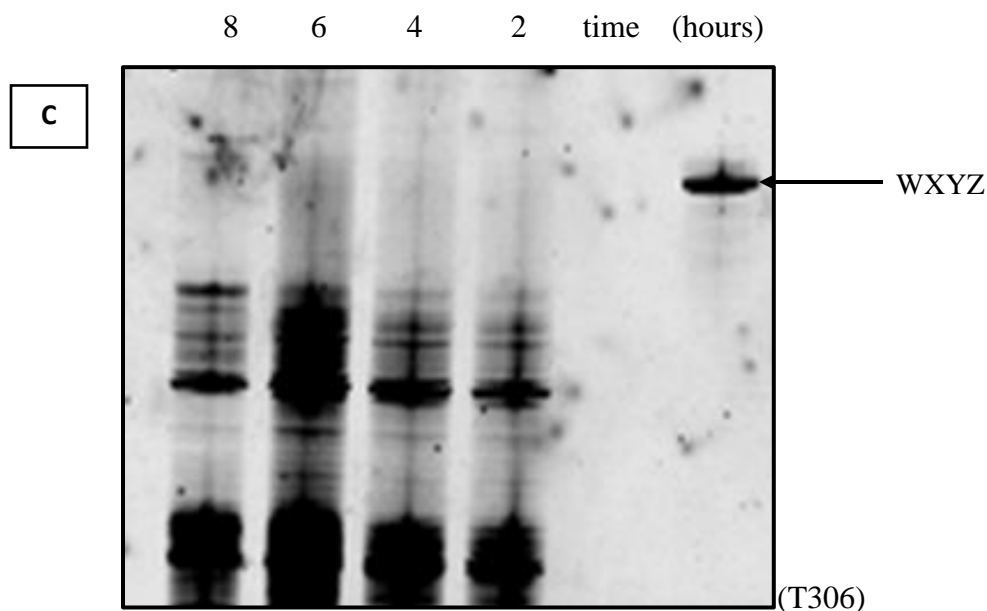


III. Self-assembly at the GUU-AAU junction.

To this point I had tried multiple self-assembly reactions in the **X**, **Y**, and **Z** fragments. The **W** fragment is the longest of all four fragments. Therefore, I decided to pick another new breaking location, where the complementary tag sequence is located in the **W** fragment. I first tried a two-piece self-assembly with **W₁** and **W₂XYZ** using a break at position A48A49U50. It was difficult to transcribe the **W₁** RNA fragment. Therefore, I decided to engineer the AAU complementary tag sequence to head portion of the **Z** fragment and try a different two-piece self-assembly. As shown in **Figure 18A**, we can see that the two-piece assembly is fast, as the formation of full-length product started after only a 2-hour incubation period, and it showed 40% of the reactant converted into the final product within the first two hours. Therefore, I tested the same reaction within a shorter duration of time (**Figure 18B**). As these results were very promising, I tested the four-piece self-assembly with fragments **W**, **X**, **Y**, and **Z** where AAU complementary tag sequence was engineered into the 3' end of **W**, **X**, and **Y** fragment and to the 5' head sequence. **Figure 18C** shows that the four-piece assembly with these fragments was successful and can proceed in to the five-piece assembly with **W₁+W₂+X+Y+Z**. When compared to the two-piece self-assembly, four-piece self-assembly reaction was relatively slow, and it had only 4% product yield after eight hours.

Figure 18: Self-assembly reactions with GUU-AAU combination. A) & B) Two-piece self-assembly using GUU-AAU pair (WXYZ+Z). C) Four-piece self-assembly using W, X, Y, and Z fragments.



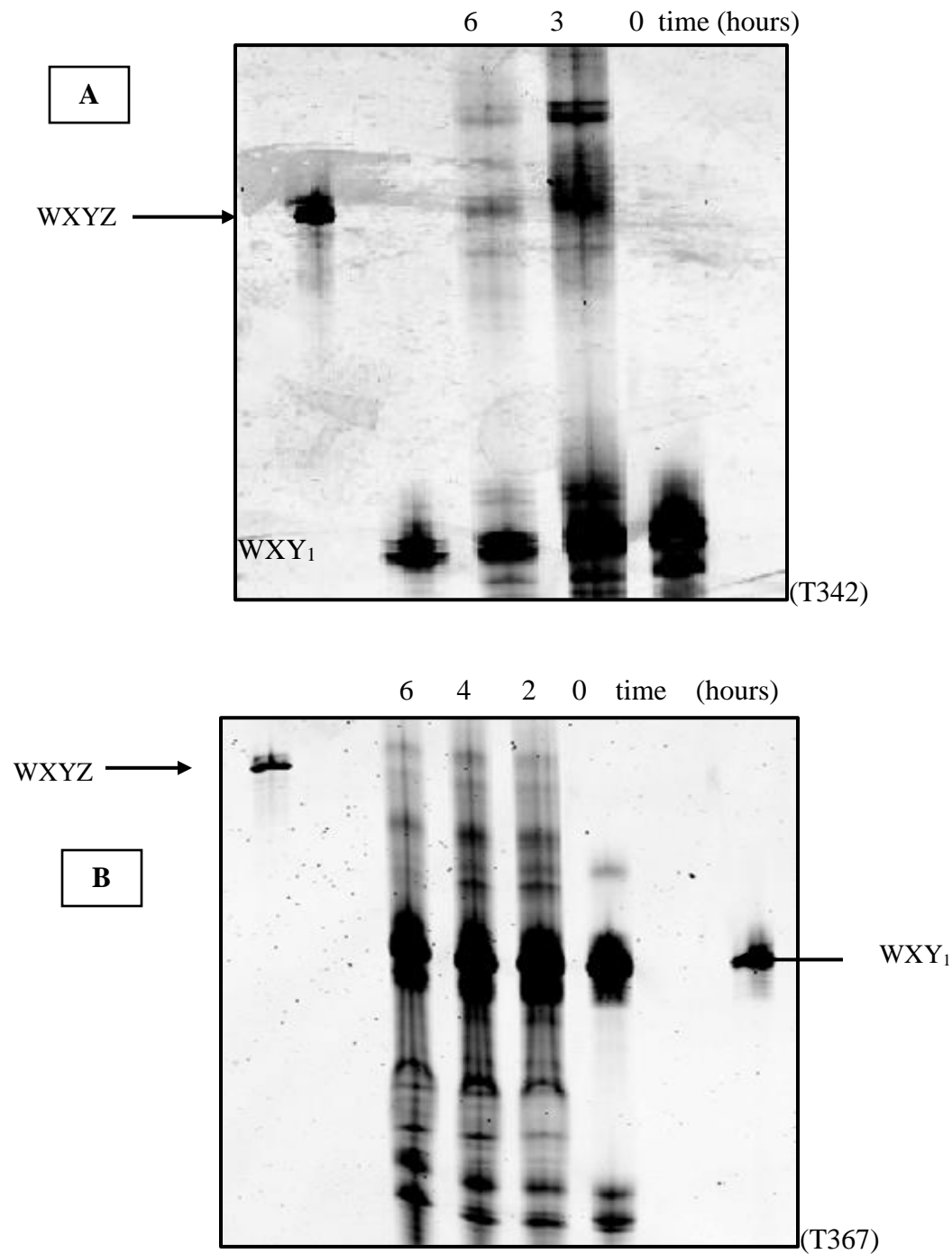


IV. Self-assembly at the GUC-GAU junction.

The complementary tag sequence for the GUC Internal Guide Sequence is GAU. This complementary tag is located in both the **W** and **Y** fragments. Structural analysis studies show that the location in the **Y** fragment is easy to access (Adams et al., 2004). Therefore, I tested a basic two-piece self-assembly reaction with RNA fragments, **WXY₁+ Y₂Z** (**Figure 19A**) at G116A117U118. The **WXYZ** product yield at the end of first three hours was 40%. Then a three-piece self-assembly reaction was tested using **WXY₁+Y₂+Z** RNA fragments. (**Figure 19B**) and it gave 6% of final product yield after two hours. However, when I tested a four-piece assembly with **W, XY₁, Y₂, and Z** RNA fragments, it was not a successful experiment, as no full-length **WXYZ** was observed after seven hours.

Figure 19: Self-assembly reactions with GUC-GAU combination. With the GUC-GAU combination, A) two-piece self-assembly reaction was tested with **WXY₁+Y₂Z** fragments.

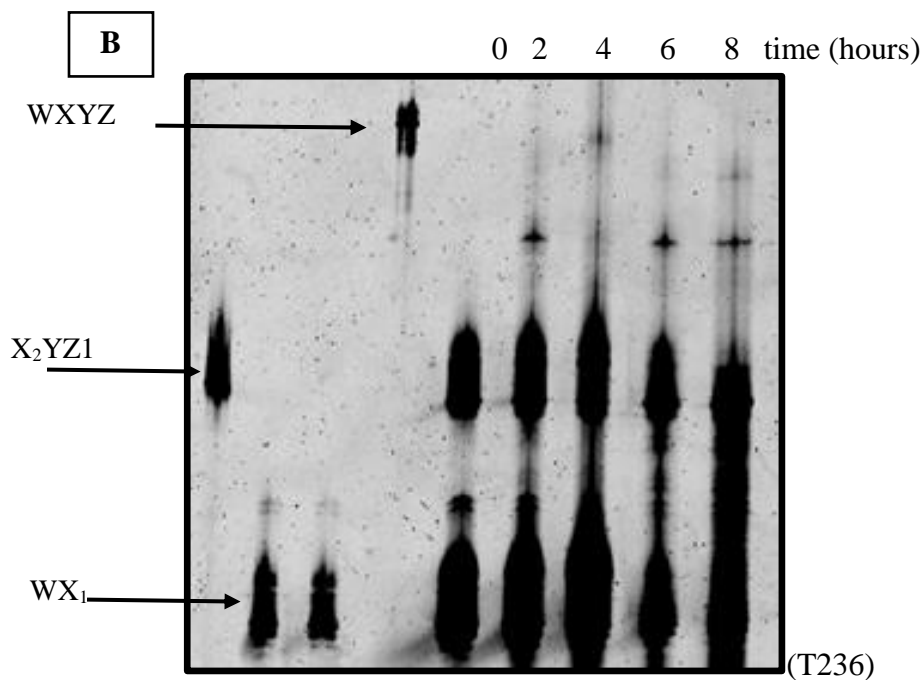
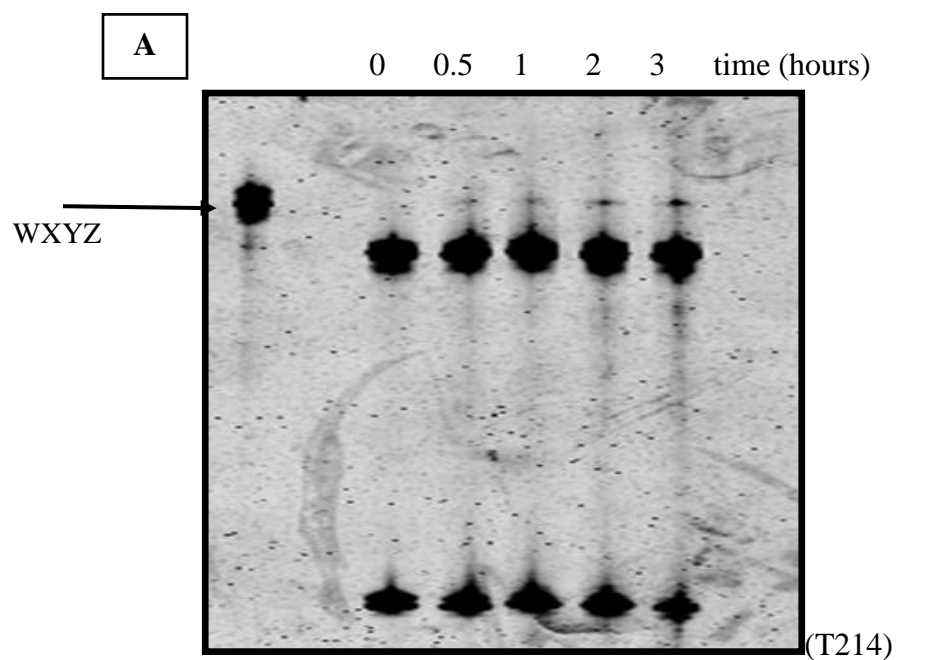
B) Three-piece self-assembly using RNA fragments **WXY₁, Y₂, and Z**.



V. Self-assembly at the GUA-UAU junction.

The complementary tag sequence UAU can be seen in both the **X** and **Z** fragments. I tested two-piece self-assembly with **WXYZ₁** and **Z₂** RNA fragments (**Figure 20A**) using the naturally occurring UAU at position at U160A161U162. Then with this new breaking location, I only observed successful assembly up to three-piece system (**Figure 20B**). Both reactions produced very poor final product yields of 2% after four hours.

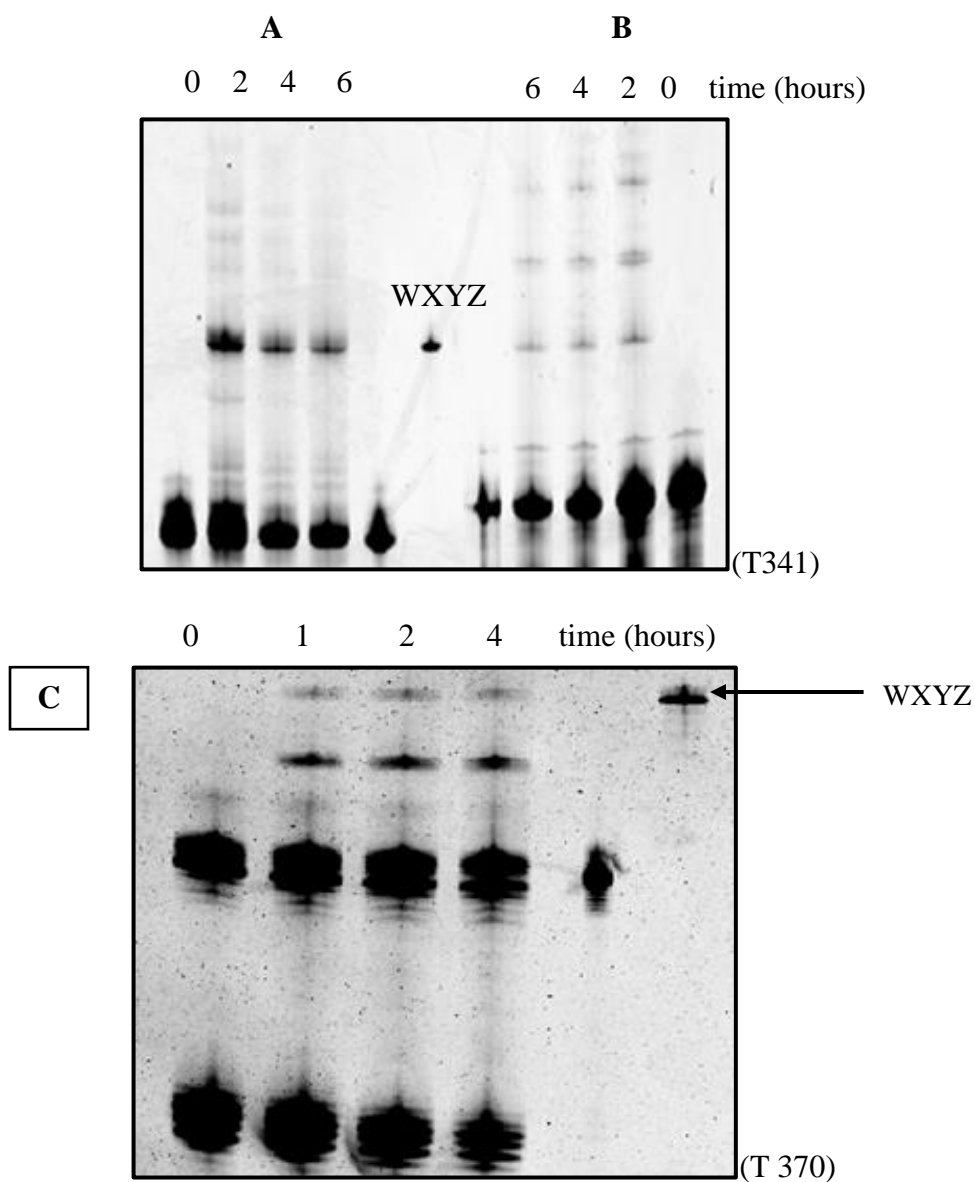
Figure 20: Self-assembly reactions with GUA-UAU combination. Using GUA-UAU pair as the breaking location, **A)** Two-piece and (WXYZ1+Z2) **B)** Three-piece self-assembly (WX1+X2YZ1+Z2) reactions were tested.

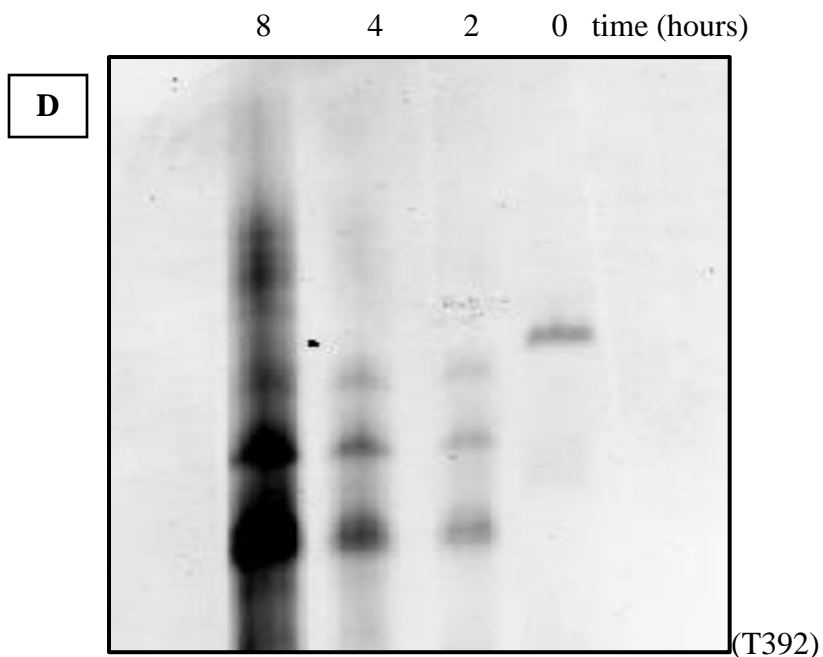


VI. Self-assembly at the GGG-CCU junction.

In this IGS-tag pair, a CCU tag is in both the **W** and **Y** fragments at positions CCU62 and CCU149. Therefore, two separate binary self-assembly reactions were carried out. As shown in **Figure 19A** and **B**, both two-piece self-assembly reactions were successful. When I tested the self-assembly between **W₁** and **W₂XYZ** RNA fragments (**Figure 21B**), the reaction gave 12% product yield after six hours. Then the second two-piece self-assembly was with **WXY₁** and **Y₂Z** RNA fragments which gave only a 3% of final product yield after 6 hours (**Figure 21A**). After considering the length of the RNA fragments that are required for the next three-piece assembly reaction, I tested only the junction in **W** fragment. Therefore, subsequent three-piece self-assembly reaction was carried out using the **W₁**, **W₂X** and **YZ** RNA fragments. As shown in **Figure 21C** this was successful and produced 7% of final product yield after four hours. I then tested the four-piece self-assembly reaction with RNA fragments; **W₁**, **W₂X**, **Y₁**, and **Y₂Z**. As shown in **Figure 21D** this four-piece self-assembly was not successful.

Figure 21: Self-assembly reactions with GGG-CCU combination. The IGS/tag pair GGG-CCU was tested. **A)** Two-piece self-assembly with fragments **WXY₁** and **Y₂Z** using **B)** Two-piece self-assembly at the junction in **W** fragments using with **W₁** and **W₂XYZ**. **C)** Three-piece self-assembly at the junction in **W** fragment with **W₁+W₂X+YZ**. **D)** Four-piece self-assembly with fragments **W₁, W₂X, Y₁, and Y₂Z**.





VIII. Self-assembly at the GGC-GCU junction

I next tested the GGC-GCU pair. This GCU tag sequence was naturally located within **X** and **Z** fragments at positions G99C100U101 and G158C159C160. After considering the locations, I decided to test the binary self-assembly reaction using **WX₁** and **X₂YZ** RNA fragments (**Figure 22A**). The final product yield of this two-piece self-assembly was 9% after six hours.

IX. Self-assembly at the GAU-AUU junction

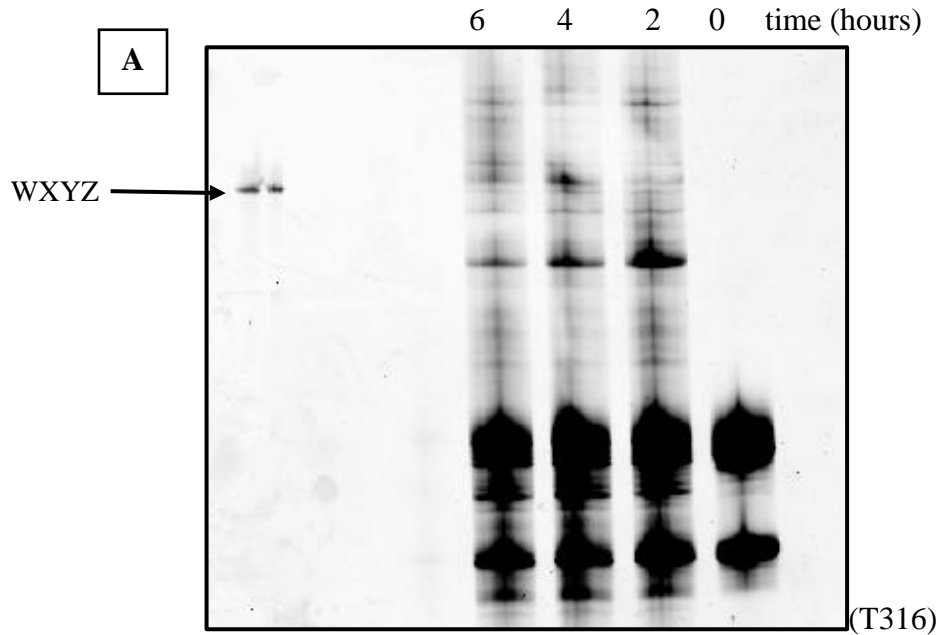
A binary self-assembly reaction was only performed using this IGS-tag pair. The tag AUU is naturally located at A49U50U51. The **W₁+W₂XYZ** fragments could self-assemble and

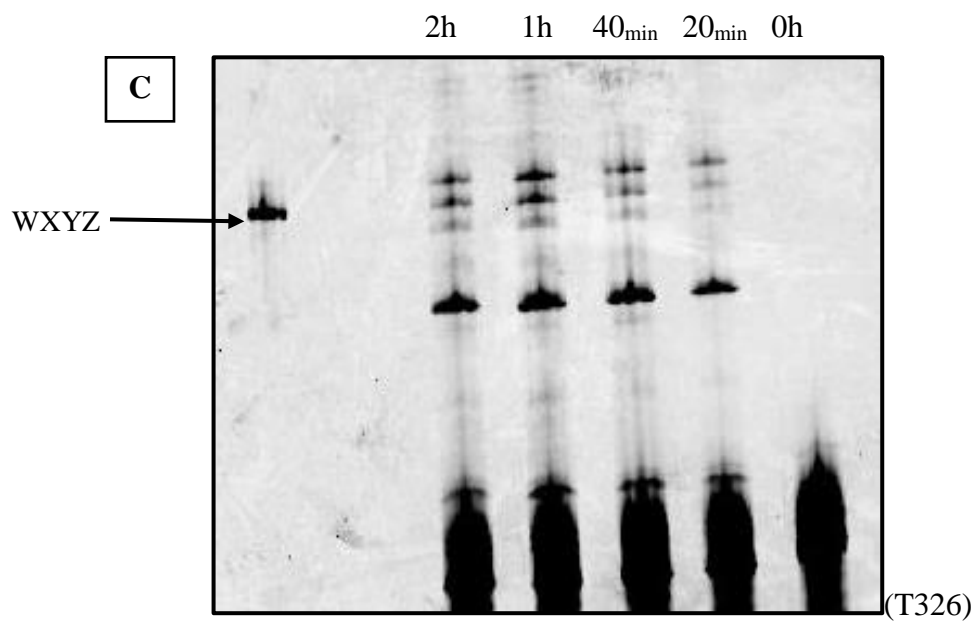
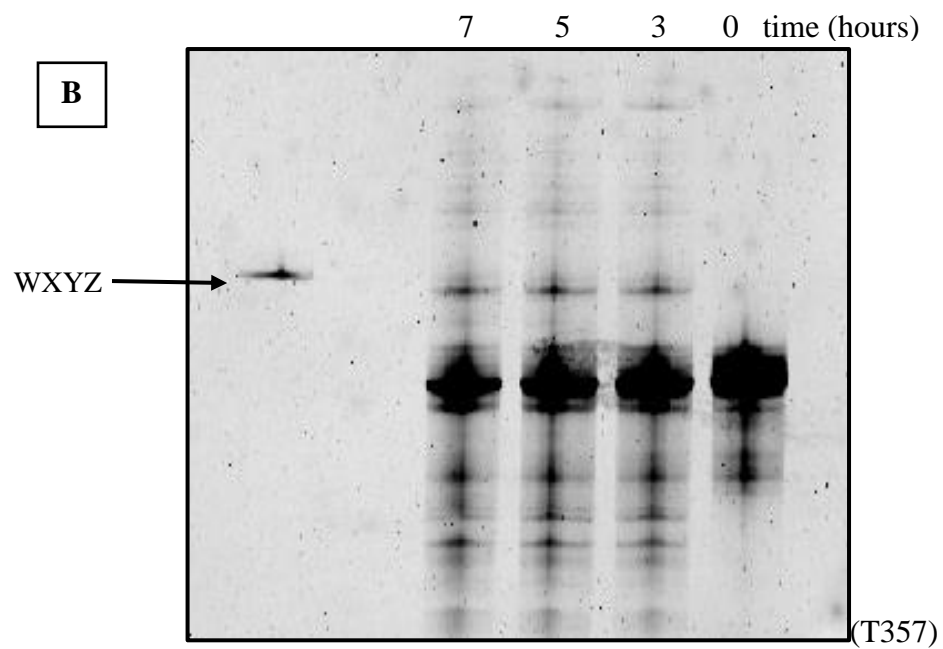
form the full-length **WXYZ** product starting from three hours through seven hours. (**Figure 22B**). This reaction had a product yield of 6% after seven hours.

X. Self-assembly at the GAG-CUU junction

Lastly, I was able to test one more binary self-assembly reaction using CUU as the new breaking location. This was located in **X** fragment at C100U101U102. Two RNA fragments were **WX₁** and **X₂YZ**. As seen in **Figure 22C** this reaction was also a successful self-assembly reaction and it gave an 8% of final product yield after two hours.

Figure 22: Self-assembly reactions with GGC-GCU, GAU-AUU, and GAG-CUU combination. (**WX₁+X₂YZ**) **B**) GAU-AUU combination (**W₁+W₂XYZ**), and **C**) GAG-CUU combination (**WX₁+X₂YZ**).





2. Combination of different self-assembly reactions

After achieving two successful five-piece self-assembly reactions, the next task was to combine these two reactions. The main purpose of this experiment was to combine two sets of self-assembly reactions which contain different IGS-tag interactions, and then observe how each variation will interact cooperatively to synthesize full-length **WXYZ** product. Here, the idea was to determine if these product ribozymes could display cooperative behavior if their IGS sequences target the assembly of other ribozymes including themselves as well (Vaidya et al., 2012). Here in the reaction mixture I combined four different 5' end **W** fragments, **X1** and **X2** contain the tag from GCG/CGU system (chapter 3, **Figure 14 C**), and **Y**, **Z1**, and **Z2** fragments contain GUG/CAU (chapter 2, **Figure 13**) in six-piece construct to test the ability of mismatched pairs to promote self-assembly.

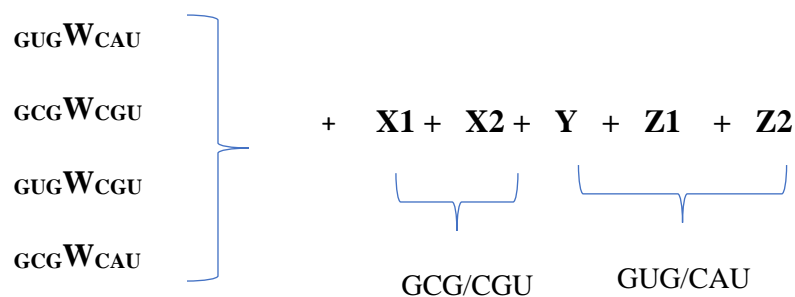
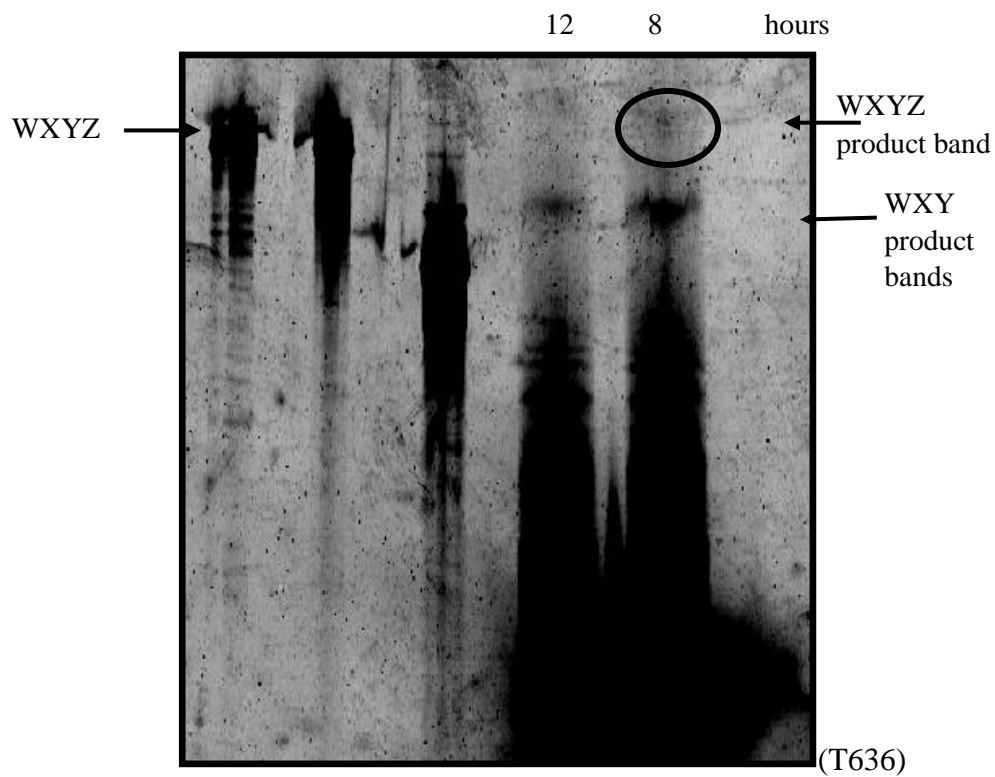


Figure 23: Combined self-assembly reaction; six-piece self-assembly. This is the polyacrylamide gel which represents the six-piece self-assembly with **W** (GUG**W**CAU, GCG**W**CGU, GUG**W**CGU, and GCG**W**CAU), **X1**, **X2**, **Y**, **Z1**, and **Z2** RNA fragments. Here we can clearly see the formation of **WXY** as a recombination product starting from eight hours and it continues to 12 hours.



Then the next task was to identify which recombination mechanism is used at each junction (Draper et al., 2008). The tag sequences CAU lies at the **Z1-Z2** recombination site and CGU sequence lies at the **X1-X2** recombination site. As I did not modify the nucleotide sequences at either **X1-X2** or **Z1-Z2** junctions, I expected the RNA fragments to recombine through R2F2 mechanism. Thus, this is also confirmed by the sequence analysis as well. Therefore, I excised the full-length band after eight hour of the six-piece self-assembly reaction, subjected it to an RT-PCT reaction specific for full-length **WXYZ**, and then cloned the DNA into an *E. coli* vector and performed Sanger nucleotide sequence analysis on six bacterial colonies. The sequence at the **W-X1**, **X1-X2**, **X2-Y**, **Y-Z1**, and **Z1-Z2** junctions shows that which recombination mechanism was used at which junction (**Table 4**). As confirmed in chapter 2 (**Table 3**) **Z1-Z2** junction have proceeded through the R2F2 mechanism. Also, as expected, 100% of the reactions at the **X1-X2** junction also have proceeded through R2F2 mechanism. The tF2 mechanism can be seen at the **W-X1** junction. However, based on the nucleotide sequences after the recombination at the **X2-Y** and **Y-Z1** it is difficult to decide which recombination mechanism was used for this reaction. Therefore, further sequence analysis with more bacterial colonies is necessary to determine the recombination mechanism at those junctions.

Table 4: Nucleotide sequences at the junctions in six-piece self-assembly. Nucleotide sequences at the junctions in three cloned full-length RNAs from six-piece self-assembly reaction. * Resulting from tF2 mechanism. + Resulting from R2F2 mechanism. At this stage it is difficult to determine which mechanism is used at **X2-Y** and **Y-Z1** junctions.

Clone	W-X1	X1-X2	X2-Y	Y-Z1	Z1-Z2
1F	GCGT*	CGT ⁺	CTG	GGCAAA	CAT ⁺
1R	GCGT*	CGT ⁺	CTG	GGAAAA	CAT ⁺
7F	GCGT*	CGT ⁺	CTG	GGCAAA	CAT ⁺
7R	GCGT*	CGT ⁺	CTG	GGCAAA	CAT ⁺
23F	GCGT*	CGT ⁺	CTG	GGCAAA	CAT ⁺
23R	-----	CGT ⁺	CTG	GGCAAA	CAT ⁺

Discussion

In this chapter I have tested seven IGS-tag combinations, which are different from the natural IGS-tag of the *Azoarcus* group-I intron. The purpose of this study was to analyze whether completely different IGS-tag combination could be used to make the covalent contiguous ribozyme from multiple short RNA fragments. Furthermore, as discussed in the Chapter 2 we saw that by using a naturally occurring tag sequence and make that locations as the new breaking point it was feasible for the short inactive RNA oligomers to go through recombination and form the full-length catalytically active ribozyme. I have been able to show that all two-, three, four, five-piece self-assembly reactions could proceed successfully using these new IGS-tag pairs. Out of all the different IGS-tag combinations that were tested GCG-CGU pair had the highest level of successful self-assembling (**Figure 12C, Table 3**). Since new mutations were not done to locate the respective tag sequence for each new IGS, it could be the reason that these reactions were very successful as shown in the Results section of this chapter. When performing the two-piece self-assembly reactions, I did have to change the nucleotide sequence inside the *Azoarcus* secondary structure sequence. The reason for this was I mutated only the IGS and located the relevant complementary tag sequence within the nucleotide sequence of the *Azoarcus* group I intron. The 5' end of the first fragment was changed according to the IGS that I engineered. Also, the relevant tag triplet was also engineered to the 3' end of this fragment. The 5' end of the second fragment contained five-nucleotide long head sequence, which has the tag triplet followed by two other nucleotides: GG. All these modifications were

done to aid the correct positioning of the nucleotides at the new junction and facilitate the nucleophilic attack during the recombination reaction.

Using these new combinations and interactions I was able show that even such interactions among completely new shorter RNA oligomers, driven mainly by non-covalent bonding strengths could then be used to show a successful five and six-piece self-assembly reactions (**Figure 11 C and Figure 23**). The successful six-piece assembly reaction (**Figure 23**) which was combination of two different IGS-tag pairs is a great example to demonstrate that even the sequence requirements are different in a prebiotic system, small RNA fragments can interact with each other through a network to produce a functional full-length product. As shown in **Table 4**, sequence analysis of the six bacterial colonies helped to determine which recombination mechanism is utilized at each of the five junctions during the six-piece self-assembly reaction. Both new junctions; **X1-X2** (from GCG-CGU five-piece assembly) and **Z1-Z2** (from GUG-CAU five-piece self-assembly) went through R2F2 mechanism, where as **W-X1** junction has gone through tF2 mechanism. As mentioned earlier, further sequence analysis of more successful bacterial colonies from the same six-piece assembly reaction is necessary to determine the recombination mechanisms at the **X2-Y** and **Y-Z1** junctions.

This chapter provides the evidence that even one of the most essential elements of a recombinase ribozyme, the IGS–tag complement pair, can withstand certain mutations away from an optimal sequence. Moreover, these systems could be used to show the

various molecular interactions among the prebiotic polymers that were available in RNA World period of early Earth.

Materials & Methods

RNA fragments

All the RNA fragments that are mentioned in this chapter have a CAU-containing short 5 nucleotide long sequence at their 5' end; 5' head (h=GGCAU) and at 3 nucleotide long tail with CAU at their 3' end. These two short sequences allow the RNA fragments to go through one of the recombination mechanisms that is shown in **Figure 4**.

RNA preparation

All the RNA fragments that are mentioned in the Results section which are portions of the self-splicing group I intron from the isoleucine pre-tRNA in the purple bacterium *Azoarcus* (Reinhold-Hurek and Shub, 1992), were prepared by *in-vitro* transcription from DNA plasmid templates and then were gel purified using 8M urea/ 8% poly acrylamide denaturing gels. Some short RNA fragments were purchased from TriLink Biotechnologies (San Diego, CA). All RNAs were resuspended in 1–10 μ M solutions in 0.1 mM EDTA. For quantification, < 0.1 μ M of the **W** containing fragment was 5'-labeled with γ [32 P]•ATP using OptiKinase (USB, Cleveland, OH).

Self-Assembly Reactions

Every single self-assembly reaction was performed in pH 7.5 reaction buffer at 48°C in a Digital Dry Bath. First, all RNA fragments and required amount of water were added to a 0.6 mL Eppendorf tube. The final RNA concentration in the reaction mixture was 2 µM. Then this RNA mixture was incubated for 3 minutes at 85°C. After that 500 mM MgCl₂ and 150 mM EPPS buffer were added. Lastly, the reaction tube was left at 48°C in the Digital Dry Bath for the desired length of time. At certain time intervals, aliquots of 5-10 µL from the reaction mixture were removed and quenched with a quench solution, which contains 500 mM EDTA and 2X Polyacrylamide Gel Loading Dye (2X loading dye containing formamide and bromophenol blue). Samples were loaded on an 8% polyacrylamide/ 8 M urea gels. Radiolabeled gels were visualized using the Typhoon Trio variable mode imager using the phosphoimaging settings and the non -radiolabeled gels were stained with SYBR Green II solution and visualized on the Typhoon Trio+ variable mode imager using fluorescence settings and products were quantified using Image-Quant software. Percent Yields were calculated as $(\text{Reacted \%} = [\text{Reacted} / (\text{Reacted} + \text{Unreacted})] * 100)$.

As described in the methods and materials section in chapter 2, the 8-hour **WXYZ** product RNA in **Figure 21** was purified, reverse transcribed and amplified using PCR for genotyping (described in chapter 2).

Chapter 4
Analysis of Effect of Evaporation and Re-hydration
on RNA Self-Assembly

Introduction

The first molecules of life emerged around 4 billion years ago under early Earth's conditions such as the composition of the Earth's crust and the atmosphere at that time. These first prebiotic molecules such as nucleic acids and amino acids must have formed through continuous physicochemical process which helped these molecules to achieve greater complexity through fluctuations of the physical parameters. Various scenarios have been used to study the beginning of life on early Earth. A fundamental problem in origins of life research is how the first polymers with the properties of nucleic acids were synthesized and incorporated into living systems on the prebiotic Earth. Therefore, to understand the origin of living systems under prebiotic conditions it is important to explore the principal possibilities of how molecules combine to produce more complicated functional units. The process of self-assembly is a necessity under proper environmental conditions. A suitable site for the synthesis of life must have had the necessary driving forces for the synthetic reactions. Moreover, there must have been chemical processes by which one or many organic compounds were concentrated so that they can undergo physical and chemical interactions.

Mineral surfaces offer an interface that help organic solutes from aqueous solutions to adsorb on to and thereby concentrate them. Ferris and co-workers have experimented the ability of montmorillonite surface to organize and catalyze the synthesis of RNA from activated mononucleotides. RNA polymers up to 50 nucleotides in length can be produced by this reaction (Ferris et al.,1996). A possible pathway for early prebiotic oligonucleotide formation may involve synthesis of random polymers which include 2',5' and 3',5' phosphodiester bonds (Usher,1977). Phosphorylation of prebiotic molecules under aqueous conditions was a critical step in the origins of life. A recent study demonstrated that diamidophosphate (DAP), a plausible prebiotic compound produced from trimetaphosphate, efficiently (amido)phosphorylates a wide variety of prebiotic building blocks such as nucleotides, amino acids, and lipid precursor molecules conditions, without the need for a condensing agent (Gibard et al., 2018). Another study showed that the activated pyrimidine ribonucleotides can be formed in a short sequence that bypasses free ribose and the nucleobases, and instead proceeds through arabinose amino-oxazoline and anhydronucleoside intermediates (Powner et al., 2009).

An important question guiding research on the origin of life concerns the environmental conditions where primordial molecular systems which have life like properties first appeared on the early Earth. A suitable site requires liquid water, a source of organic compounds, a source of energy to drive polymerization reactions, and a process by which the compounds were sufficiently concentrated to undergo physical and chemical interactions. A site that promotes self-assembly of organic solutes with an aqueous

environment containing various ionic solutes, at an intermediate temperature range and neutral pH ranges, in which cyclic concentration of the solutes can occur by transient wet-dry intervals. Another important aspect is by combining different chemical synthetic processes it might be possible and the resultant complex structures can then increase the efficiency of the synthesis process which would be useful for the emergence of prebiotic chemical systems in aqueous environments toward an RNA world period.

Geothermal sites with mixtures of organic solutes and phosphates mixed with high-temperature water in clay pools have been studied. Analysis of kaoline clay showed that the organic molecules were adsorbed to the mineral surfaces at the acidic pH first (Deamer et al., 2006). When considering the synthesis of primordial molecules, the optimum temperature plays a crucial role by facilitating the favorable physiochemical reaction conditions. An *in vitro* evolution of RNA catalysts directly in the water-ice medium can yield RNA polymerase ribozymes which are specifically adapted to sub-zero temperatures (Attwater et al., 2013). Here, these catalysts can synthesize RNA in ice at temperatures as low as -19 °C. An efficient way of assembling RNA polymerase ribozyme from mixtures of RNA oligomers no longer than 30 nucleotides has been shown (Mutschler et al., 2015) using freeze thaw cycles. Where the freeze-thaw cycles can be used to drive the self-assembly of RNA polymerase ribozyme. This was one of the ways to bridge the gap between the primitive pools of short RNA oligomers and complex ribozyme structures. Another study inferred that RNA-like polymers can be synthesized non-enzymatically from 5'-phosphate mononucleosides in salt environments (Da Silva et al., 2015). Here,

they showed that RNA-like polymers can be synthesized non-enzymatically from 5'-phosphate mononucleosides in salty environments. The synthesis of phosphodiester bonds was driven by the chemical potential made available in the fluctuating hydrated and anhydrous conditions of hydrothermal fields associated with volcanic lands.

Life may have originated in a hydrothermal milieu on land rather than in a deep-sea environment. Hydrothermal fields have fluctuating environments where dehydration and rehydration cycles occur in small ponds-like environments. Hydration and dehydration cycles would have been occurred on early Earth when volcanic land masses emerged from the ocean. Therefore, these hydrothermal fields provide ideal environments for prebiotic polymer synthesis. Several recent studies used hydration-dehydration cycles to show that the condensation products of mononucleotides and hydrolysis of their polymers (Damer and Deamer, 2015). During this stage on early Earth, hydrothermal fields were present with small aqueous pools which were undergoing evaporation and refilling at higher temperatures. One study which used the simulation of hydration and dehydration cycles showed that polymers which resemble the properties of nucleic acids can be synthesized using these simulation cycles, even in the absence of an enzyme or activated substrates (DeGuzman et al., 2014).

Geothermal field with fluctuating wet-dry cycles are also excellent prebiotic sites for appearance of life-like molecules on the early Earth. These wet-dry cycles may have driven chemical reactions which allow the emergence of life-like prebiotic molecules. Wet-dry cycles can fluctuate the chemical environment that is required for polymerization reactions.

These cycles can act as the possible free energy source for these reactions. The physical parameters established by these wet-dry cycles might have allowed for the formation of self-replicating RNA as the fundamental biological molecule on the early Earth. Studies have shown that geothermal fields could provide the right environmental conditions for the synthesis of RNA monomers by continuous synthesis by changing physical parameters such as temperature, pH, and concentration driven by wet-dry cycles (Becker et al., 2018). These fluctuating environments that were caused by continuous wet-dry cycles might have favored the formation of reaction intermediates that are directly used for the next step in the synthesis process. Therefore, a continuous synthesis of complex prebiotic polymers was established.

A model that shows the molecular systems driven by hydration and dehydration cycles in hydrothermal sites could undergo chemical evolution in dehydrated films on mineral surfaces which then will be followed by encapsulation and selection in a hydrated phase (Ross and Deamer, 2016). The dehydration phase helps to increase the concentration of the solute mixtures. Also, during the dehydration phase, solutes and amphiphiles accumulate on mineral surfaces to form highly concentrated films. These environmental conditions help to organize and concentrate monomers and help polymerization reactions to occur through reduced water activity in the dehydrated phase because the chemical potential required to drive condensation is available when the water activity is reduced. Long strands of RNA-like polymers can be synthesized using lipid microenvironments through condensation reactions when the reactants are subjected to several rounds of

dehydration and rehydration (Rajamani et al., 2008). This shows that fluctuating environments were common on the prebiotic Earth and led to polymerization of prebiotic polymers. These studies help to understand the range of possible environments for the origin of life. All these research experiments suggest that RNA building blocks can indeed be formed in a prebiotically plausible geochemical environment.

The type of the salts presents and their concentrations in the prebiotic environment play a major role in the molecular synthesis on the early Earth. Metal ions are required for the assembly of biologically active RNA and RNA-protein complexed, including group I and group II introns, spliceosomal RNAs, and rRNA. Structural studies of these bound metal ions have shown two main roles: most of the metal ions are used for the structural stability in order to stabilize the active site then a few of the metal ions are used in the catalytic process of the ribozyme (Hanna and Doudna, 2000).

The structure and activity of RNA depend on the type and concentration of counterion in the solution. The cations in the solution and negatively charged phosphate groups in the nucleic acids form strong electrostatic interactions. As a result of these interactions the hydration shells surrounding the cation and the RNA remain intact. Magnesium ions play a crucial role in reducing repulsions and stabilizing the folded conformation of the RNA. Magnesium ions interact with RNA in several different ways. This includes diffuse including, outer- sphere, and inner-sphere complexes (Misra and Draper, 1999). Studies

have shown that Mg^{2+} is specifically required to maintain the active conformation of the catalytic core of the *Azoarcus* ribozyme. And the few Mg^{2+} ions that bind near the active site residues can induce a structural change in the catalytic core of the ribozyme that facilitates the self-splicing activity (Rangan and Woodson, 2003).

Results

In this chapter, I investigated multiple self-assembly reactions using the *Azoarcus* group I intron as the model system. Following an observation that the evaporation and condensation has a direct effect on the self-assembly reaction, I conducted self-assembly experiments with many small RNA fragments including the main five-piece self-assembly reaction with the **W**, **X**, **Y**, **Z1**, and **Z2**, I continued using the heat block and analyzed how evaporation and rehydration can improve the final product yield. First, I analyzed the effect of evaporation and condensation of the aqueous medium on the self-assembly reaction. Then I tested the effect of different MgCl₂ concentrations on the self-assembly reaction. Using these experiments, I was able to optimize the reaction conditions for self-assembly reactions using short RNA oligomers including the five-piece self-assembly reactions described in chapters 2 and 3.

To identify if there is a direct effect from the evaporation on the five-piece self-assembly reaction, first I tested a simple two-piece self-assembly reaction between **WX1** and **X2YZ** RNA fragments in a thermal cycler machine and in a digital dry bath. As shown in **Figure 25** both of these reactions were carried out for 16 hours. Although a full-length **WXYZ** product was not observed on any of these gels, an intermediate product (circled on **Figure 25**) was observed in panel **A**. Whereas such intermediate product was not seen in panel **B**. Due to the metal heated lid in the thermal cycler machine, it is not possible for the aqueous medium to evaporate but when in the heat block the aqueous medium evaporates and

condensates under the lid of the Eppendorf tube. This clearly indicates that due to this evaporation self-assembly reaction was possible when it was carried out in a heat block.

After that I tested the effect of evaporation and rehydration on the five-piece self-assembly reaction by spinning the reaction volume down via a centrifuge at different time points. It was clear that letting the aqueous medium evaporate for a longer period and then centrifuging for few seconds (so that the condensed water can again solvate in the reaction) has a direct impact on the self-assembly reaction. Spinning down at different times such as 6 hours (tube **A**), 8 hours (tube **B**), or at 9 hours (tube **C**) gives the reaction mixture a different time range for the evaporation. **Figure 26** shows that the full-length product forms regardless how long the reaction was subjected for evaporation. Therefore, I wanted to test what is the optimized time length for self-assembly yields. Different time points were picked to test: **tube A** was spun down after 6 hours of evaporation, **tube B** was spun down after 8 hours, and the **tube C** was spun down after 9 hours. By only letting the reaction to evaporate only 9 hours and then immediately spinning down gave a 2% product yield whereas tube **A** and **B** gave very low product yields after 6 hours and after 8 hours.

In the next step of experiments, I tested the efficiency of self-assembly with different MgCl_2 concentrations. I tested the effect of coupling the evaporation-rehydration technique with different lower MgCl_2 concentrations. As shown in **Figure 27**, a full-length product formed with 50 mM MgCl_2 as the starting concentration and I was not able to detect any

major bands for **WXYZ** product when I used 75 mM or 100 mM as the starting MgCl_2 concentration. At 50 mM MgCl_2 I was able to see 1.2% product yield after 12 hours which was higher than the % product yields from 75 mM and 100 mM MgCl_2 . Therefore, I compared the same experiment with 100 mM and 50 mM MgCl_2 (**Figure 28**). This experiment showed that more full-length product produced in 50 mM MgCl_2 solution. This showed that lowering the Mg^{2+} concentration, under evaporation-rehydration experimental conditions, favors the self-assembly reaction when compared with 100 mM MgCl_2 . Then the Mg^{2+} concentration was lowered to 25 mM and see whether the five-piece self-assembly reaction is feasible. As shown in **Figure 29** this reaction with 30 μL was incubated in a closed 0.65 mL tube for nine hours in a heat block, after which more than 90% of the liquid had evaporated and condensed under the lid of the tube. At that point, tube was centrifuged to bring the liquid back to the bottom of the tube, and the reaction was continued in the heat block for another three hours. I was able to see that even with 25 mM MgCl_2 , five-piece self-assembly reaction was possible with a 5.0% product yield (**Figure 29**).

After that I optimized the total initial reaction volume for the five-piece self-assembly reaction (Figure 30). I performed a five-piece self-assembly reaction in two separate tubes; tube A and B. Where tube **A** had 40 μL and tube **B** had 30 μL as the initial volumes. When the reactions were performed for 16 hours and the product yields were calculated, tube **B** which had 30 μL as the starting volume gave a higher product yield (3.5%) when compared to tube **A**'s product yield (1.8%).

Therefore, after optimizing initial MgCl_2 concentration, the length of evaporation time, and initial total reaction volume, I decided to continue the future five- piece self-assembly reactions with **W**, **X**, **Y**, **Z1** and **Z2** fragments using 25 Mm MgCl_2 , 30 μL initial volume and let the reaction run up to 9 hours and then give an immediate centrifuge to rehydrate (**Figure 24**). After that I measured RNA yields and calculated the change of concentration of the reactant RNA fragments and Mg^{2+} due to the evaporation and spinning down during the total reaction period. For this I measured the remaining aqueous volume at the bottom of the Eppendorf tube after every hour. I then used these readings then to calculate the final RNA concentration and MgCl_2 concentration at every hour (**Tables 5, 6, 7, and 8**).

Figure 24: Illustration of testing the effect of evaporation and rehydration. Schematic for testing the effect of evaporation and immediate centrifugation. The 1.7 mL Eppendorf tube containing the reaction mixture was incubated for the desired amount of time (hours) in a heat block at 48°C . And then immediately was subjected to quick spun down after six, eight or nine hours to rehydrate the reaction mixture with the condensed liquid that was collected underneath the lid of the Eppendorf tube.

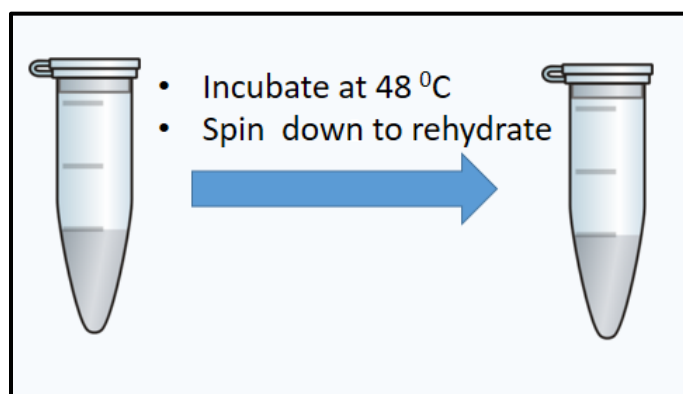


Figure 25: Self-assembly reactions using a heat block and thermal-cycler. A binary self-assembly reaction was performed by incubating **WX1** and **X₂YZ** RNA fragments at 48°C. Panel **A** was tested in a digital dry bath (heat block) and panel **B** was tested in a thermal cycler machine. A major band for an intermediate product was observed in panel A whereas there are no intermediate product bands appeared in panel B.

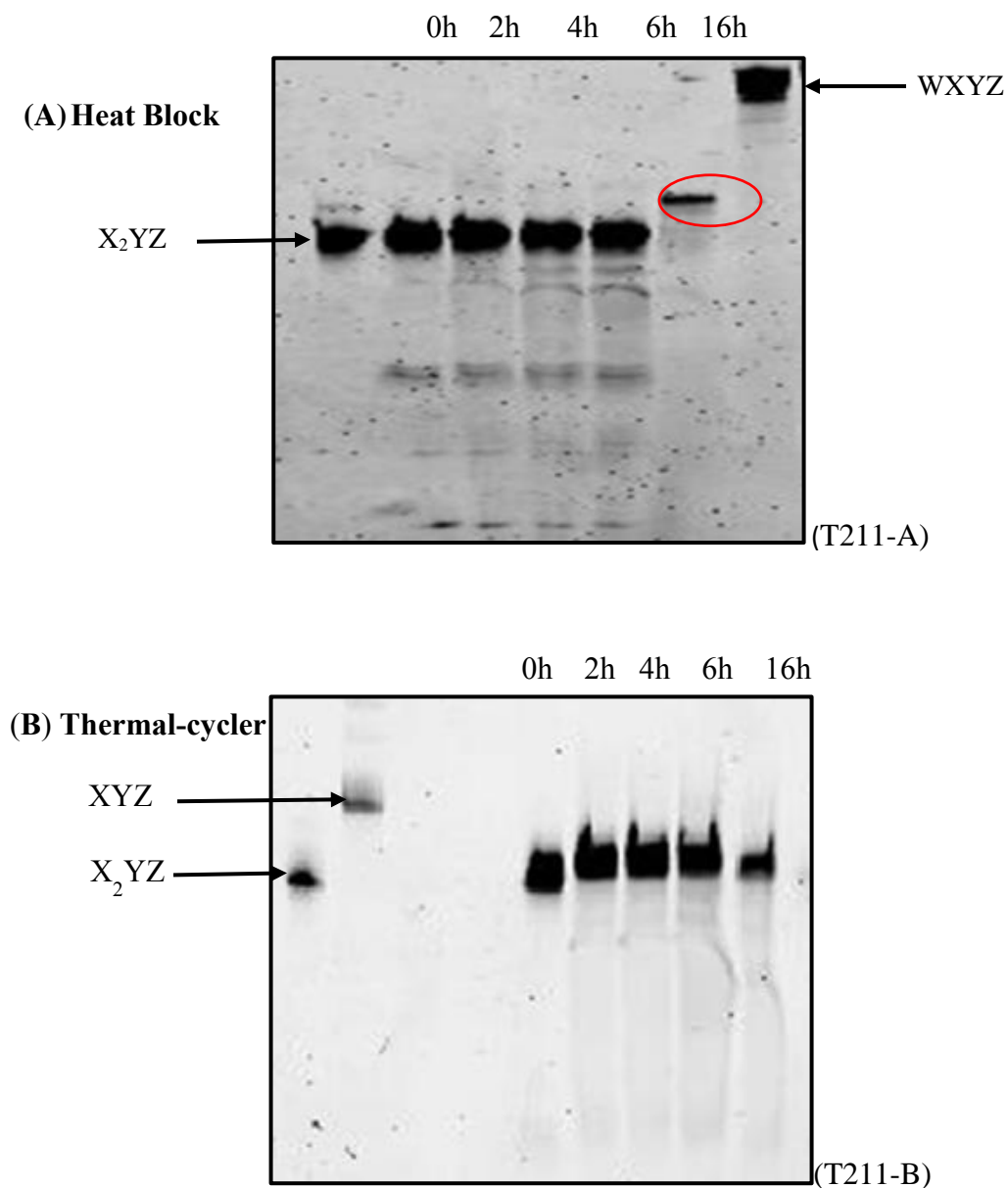


Figure 26: Effect of the length of the evaporation time on five-piece self-assembly. A denaturing gel of the five-piece self-assembly, depending on the time of intermediate centrifugation. The five-piece self-assembly reactions were performed by incubating **W**, **X**, **Y**, **Z1**, and **Z2** RNA fragments at 48°C. Three tubes were subjected to different length of evaporation. The full-length WXYZ can be seen in all three reaction tubes; **A**, **B**, and **C**. However, Tube **C** shows the highest intensity of the product band.

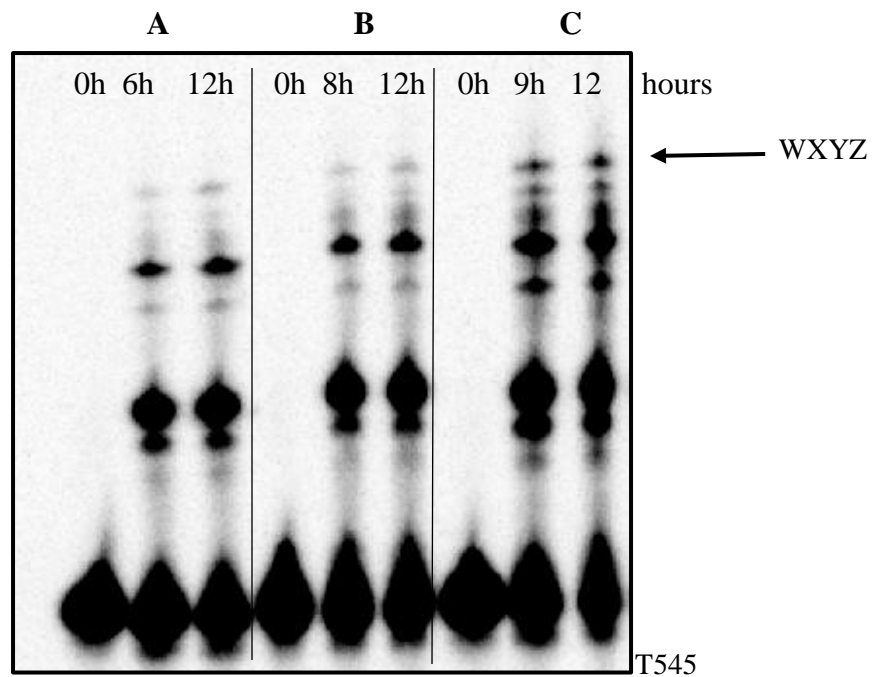


Figure 27: Effect of different Mg^{2+} concentration on five-piece self-assembly. Five-piece self-assembly reaction was tested at different Mg^{2+} concentrations at 48°C. A band was observed on the gel when 50 mM $MgCl_2$ was used.

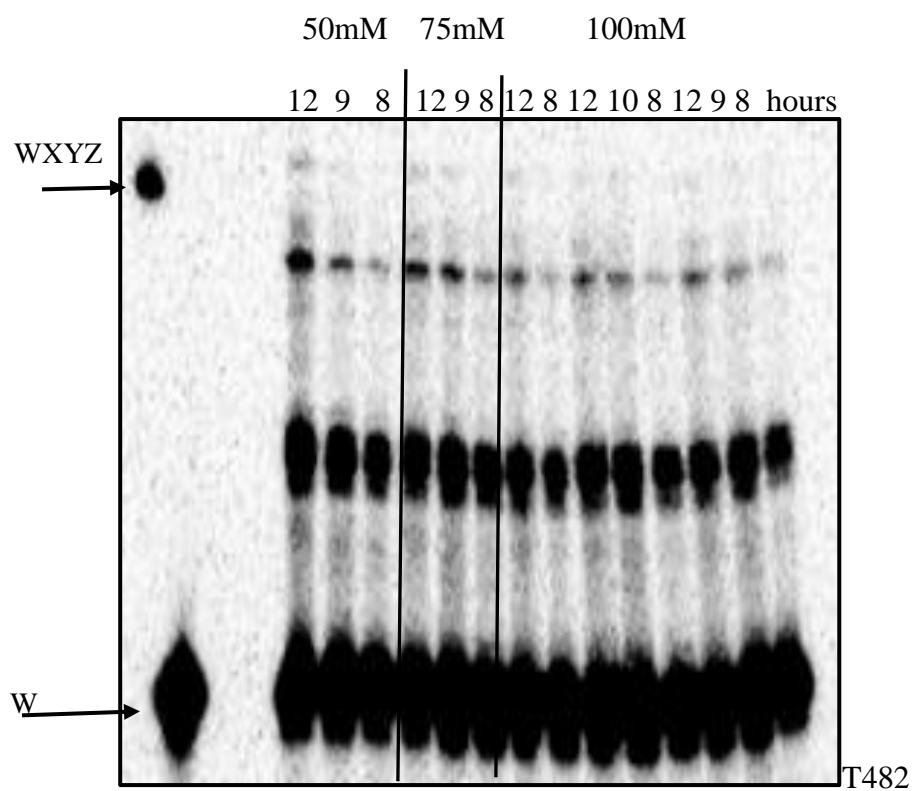


Figure 28: Effect of 50 mM and 100 mM Mg^{2+} concentration on five-piece self-assembly. Comparison of five-piece self-assembly between (A) 50 mM and (B) 100 mM $MgCl_2$. Product bands of **WXYZ** can be seen from 12 to 17 hours in panel B whereas no product bands for **WXYZ** in panel A.

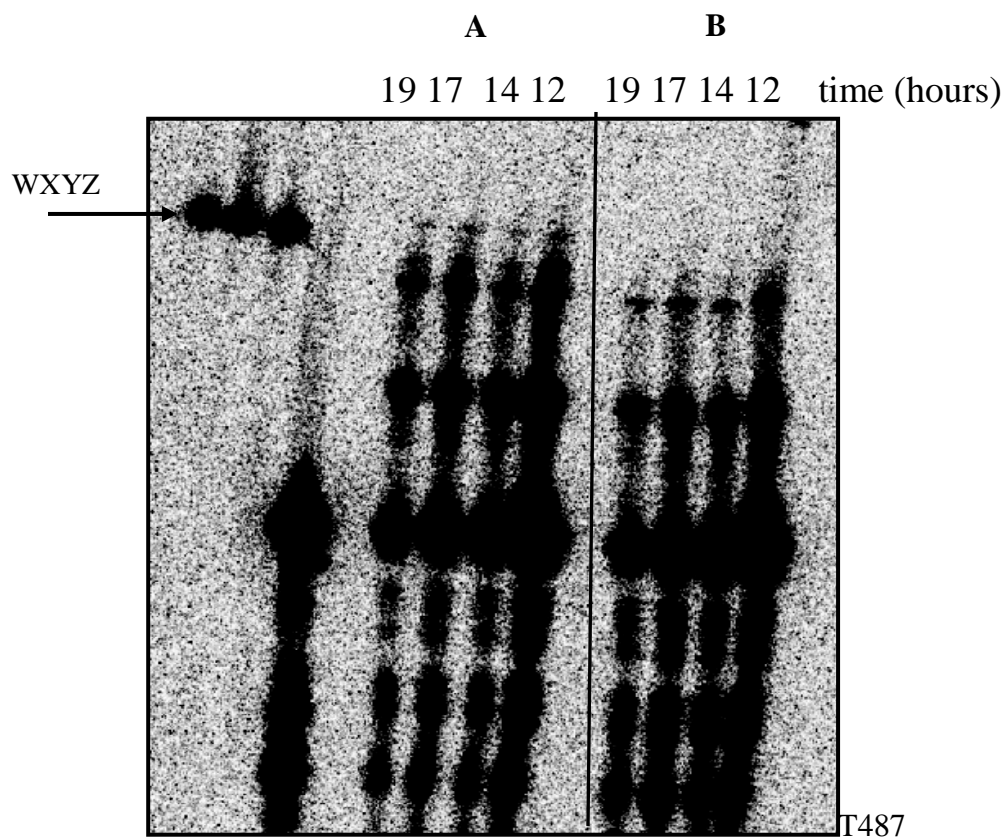


Figure 29: Effect of 25 mM and 100 mM Mg^{2+} concentration on five-piece self-assembly. Comparison of RNA product yields between (A) 25 mM and (B) 100 mM Mg^{2+} concentrations. A five-piece assembly reaction was performed at 48°C. A sharp band for **WXYZ** full-length was observed when 25 mM $MgCl_2$ was used.

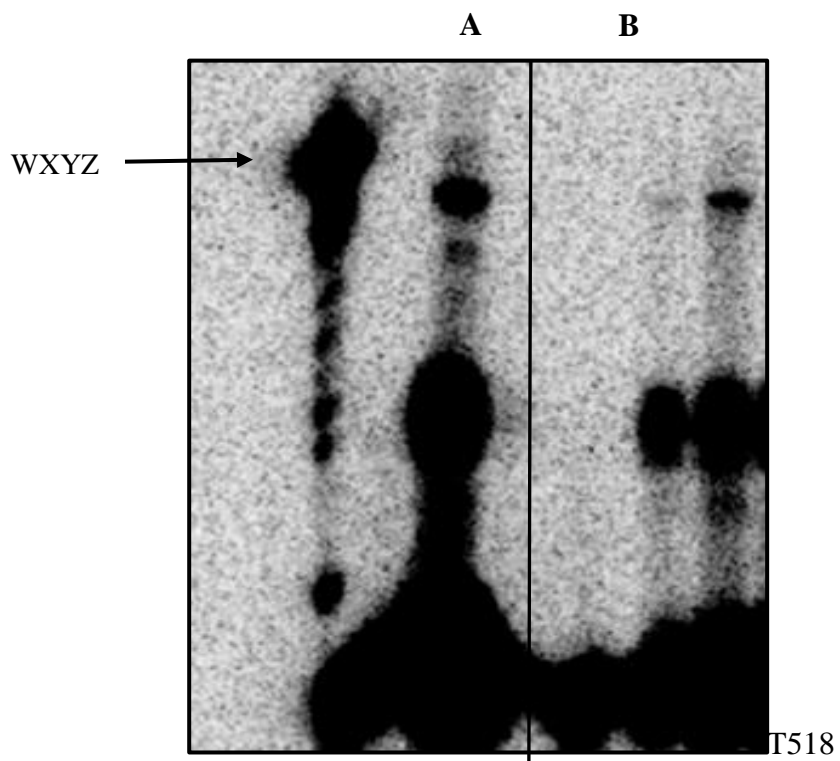


Figure 30: Effect of using different total reaction volumes on five-piece self-assembly. Testing the optimum total reaction volume for the five-piece self-assembly reaction. Five-piece self-assembly reaction was carried out in two different tubes which contain different total volumes as the initial reaction volume; tube **A** (40 μ L) and tube **B** (30 μ L).

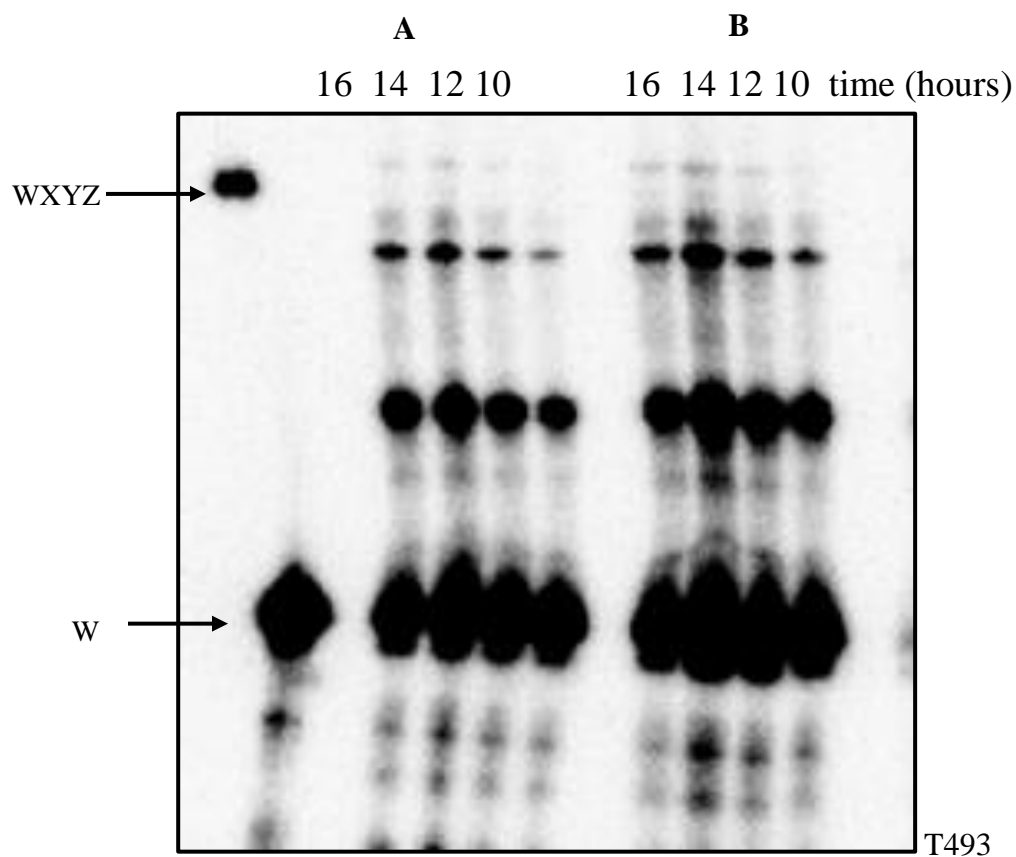


Table 5: The change of total reaction volume, RNA and MgCl₂ concentrations after six hours of incubation. The change of total reaction volume, RNA and MgCl₂ concentration with time when the reaction tube was subjected for evaporation until six hours and then spun down afterwards.

Time	Volume(μ L)	RNA concentration (μ M)	MgCl ₂ concentration (μ M)
0	25	2.4	30
1	21	2.8	35.7
2	19	3.1	39.4
3	17	3.5	44
4	13	4.6	57.5
5	11	5.4	68
6	9	6.6	83.1
7	11	5.4	68
8	9.5	5.2	78.7
9	7	8.4	106.8
10	5.5	10.7	136
11	6	9.8	124.7
12	4.5	13	166.2

Table 6: The change of total reaction volume, RNA and MgCl₂ concentrations after eight hours of incubation. Here the reaction tube was subjected for evaporation until eight hours and then spun down afterwards.

Time	Volume(μ L)	RNA concentration (μ M)	MgCl ₂ concentration (μ M)
0	25	2.4	30
1	20.5	2.9	36.5
2	18	3.3	41.6
3	14	4.2	53.4
4	12	4.9	62.3
5	9.9	5.9	75.5
6	8	7.3	93.4
7	5.5	10.6	135.8
8	4	14.6	186.7
9	10	5.8	74.7
10	8	7.2	93.4
11	6	9.6	124.5
12	3	19.2	249

Table 7: The change of total reaction volume, RNA and MgCl₂ concentrations after nine hours of incubation. Here the reaction tube was subjected for evaporation until nine hours and then spun down afterwards.

Time (h)	Volume left (μL)	RNA concentration (μM)	MgCl ₂ concentration (μM)
0	25	2.4	30
1	22.25	2.7	33.3
2	18.6	3.2	39.7
3	16.0	3.7	46.1
4	13.2	4.5	56.3
5	10.9	5.4	67.6
6	8.8	6.6	83.7
7	6.8	8.5	108.3
8	5.0	11.7	150.2
9	3.2	17.9	229.9
10	10	5.7	73.5
11	4	14.2	183.7
12	4	14.2	183.7

Table 8: Percent molar yields of the five-piece self-assembly, depending on total reaction time and time of intermediate centrifugation.

time (hours)	Spin after 6 hours	Spin after 8 hours	Spin after 9 hours
0	0	0	0
10	0.52	0.80	3.80
12	0.99	1.30	4.60

Discussion

When a two-piece self- assembly reaction (**Figure 25**) was tested one in a thermocycler machine and the other in a heat block, both at 48⁰C, I observed an intermediate product formation in the tube that was kept inside the heat block, but not in the one in the thermocycler machine. The simple explanation for this is that the heated metal lid of the thermocycler machine prevents the condensation of the aqueous medium of the reaction mixture. On the other hand, when in a heat block there is no heated metal lid, the aqueous medium can evaporate and condensate under the lid of the tube. This will increase the concentration of the reactants: individual RNA fragments and MgCl₂. As shown in chapter 2, figure 11 that by letting the entire reaction mixture of the five-piece self-assembly system to go through evaporation until nine hours, maximum product yield of 16% of the full-length can be obtained. It was important to find out what happen to the Mg²⁺ and RNA concentration over time during the evaporation process. This effect is well illustrated in Tables 3-5 which show how these values have changed over time because of the effect of evaporation. The results reported here conclude that the optimized reaction conditions for five-piece self-assembly reactions are: 25 mM MgCl₂ concentration in a pH 7.5 buffer, total volume of 30 µL at 48⁰C. I show here that by implementing these reaction conditions a maximum product yield can be obtained through a 9-hour evaporative process in a heat-block. This has been recently published (Jayathilaka and Lehman, 2018).

It is highly likely that the prebiotic environment consisted of a variety of organic compounds and salts. Most of these compounds would fall into a global ocean to form an extremely dilute solution with all the reactants that are necessary to form prebiotic life. However, when the water in this global ocean evaporates due to heat from the sun, these diluted compounds and salts can accumulate on the surface of the large pools where they can undergo different chemical reactions, such as molecular polymerization.

These chemical reactions and molecular evolution in aqueous systems, lakes or other evaporating water bodies would have been possible on early Earth because reactants could have been concentrated and polymerized rapidly in this environment. Thus, life-like polymers could have originated near the surface due to continuous cycles of evaporation and rehydration.

Materials and Methods

RNA oligomers that are used in this chapter were from the five-piece self-assembly reactions on the **Z** fragment using GUG-CAU combination. Figure 24 is the four-piece self-assembly reactions on the **X** fragment using GUG-CAU combination.

As described in chapter 2, the RNA fragments of the *Azoarcus* ribozyme were prepared by T7 RNA polymerase run-off transcription from double-stranded DNA templates made by a PCR based synthesis techniques. All RNA fragments were gel purified using 8M Urea/ 8% Polyacrylamide denaturing gels. Some short RNA fragments were purchased from TriLink Biotechnologies (San Diego, CA). All RNAs were resuspended in 1–10 μ M solutions in 0.1 mM EDTA. For quantification, < 0.1 μ M of the **W** containing fragment was 5'-labeled with γ [32P]•ATP using OptiKinase (USB, Cleveland, OH).

Self-assembly reactions were done exactly as described previously in chapter 2, materials and methods.

Chapter 5
Techniques for Establishing an RNase- and PCR Contamination-free
Research Laboratory

Introduction

The studies of ribozyme molecules and other RNAs have been revolutionized in past years with using of various purification, amplification, and analytical techniques. One potential problem when working with RNA is the presence of contamination in extraction kits and other laboratory reagents. Therefore, the experimental studies that were done in chapter 2-4 required extreme precautions against RNase and PCR contaminations. Thus, in this chapter, I have discussed useful guidelines to mitigate the effects of contamination. This chapter mainly suggests that caution should be advised when handling RNA and awareness of this issues is important to ensure that experiments are erroneous. Therefore, this chapter can be used as a good introduction to the laboratory techniques and as a useful reference during laboratory work to minimize or avoid contamination problems.

As the trend towards more sensitive molecular biological techniques continues, it is increasingly critical to establish accurate laboratory practices and implement them. The importance of such knowledge and principles becomes useful not only when someone is planning to start up a new laboratory, but also to run a contamination-free laboratory. Practicing contamination-free laboratory methods has a huge impact on the quality and the

quantity of the research data, especially in labs using the Polymerase Chain Reaction (PCR) and/or RNA.

The Polymerase Chain Reaction (PCR) is an ingenious tool for molecular biology. During the past three decades, PCR and several other DNA/RNA amplification techniques have become useful diagnostic tools (Persing, 1991). Unfortunately, contamination of samples in a PCR can have far-reaching consequences. The sources of contaminants are diverse and sometimes very unexpected. The general strategy for preventing contamination in DNA or RNA related experiments can be divided into two parts: first, how to keep the risk of contaminations as low as possible, and second, if a contamination occurs, how to eliminate it. Prevention of DNA or RNA contamination is part of the job, but if contamination occurs, elimination is the most critical step. In general, removal of contaminant can be performed before amplification, to avoid false-positive results in the experiment and to reduce the risk of contaminating following experiments.

High quality RNA is also required for many molecular biology techniques. Obtaining high quality, intact RNA is the first and often the most critical step in performing quantitative RT-PCR, RNA sequencing, catalytic RNA assays, or many other RNA-dependent experiments. The isolation and use of purified RNA in the laboratory is complicated by the fact that RNA is significantly more labile than DNA, both in chemically and biologically. RNA is more susceptible to degradation than DNA, due to the ability of the 2' hydroxyl groups to act as nucleophile. The choice and optimization of RNA purification methods

are important for successful isolation of quality RNA and consistent performance of applications. To avoid contamination or loss of RNA, caution should be exercised to avoid protein nuclease contamination of the RNA samples during preparation and handling. The success of RNA preparation is critically dependent on the elimination of ribonuclease contamination. Protein ribonucleases are extremely stable enzymes. These RNases are resistant to metal chelating agents, and some of them, such as those in the RNase A family, can survive prolonged boiling or autoclaving (Walker and Lorsch, 2003). These enzymes generally have very high specific activity, meaning tiny amounts of contamination in an RNA sample is sufficient to destroy one's RNA. The major sources of RNase contamination in a typical laboratory include: aqueous solutions, reagents used in experiments, environmental exposure, RNases in the air, surfaces, dust, human contact with hands, skin, saliva, and sweat. Cleanliness is thus a most crucial factor for a contamination-free laboratory. With PCR and RNA, it is important to have an orderly and clean lab. With the smallest contaminant, the result of an experiment can be extremely aberrant.

Maintaining a clean and tidy laboratory will help to decrease the risk of accidents and laboratory associated infections, as well as cross contamination. Moreover, for most laboratory instruments to function well, the laboratory should be dust- and contamination-free because most of these devices can be “sinks” to distribute unwanted contamination. Great care should be taken to ensure that laboratory machines are well maintained and

reliable. Furthermore, contamination-sensitive instruments such as pipettes and microcentrifuges need to be clean routinely to ensure that they are dust and contamination free. These strict measures help to increase the performance, reliability, and lifetime of such laboratory machines and instruments. By establishing a proper quality control routine, it helps to identify and correct the errors that might occur during processing as well as recording results and data (Innis et al., 1990).

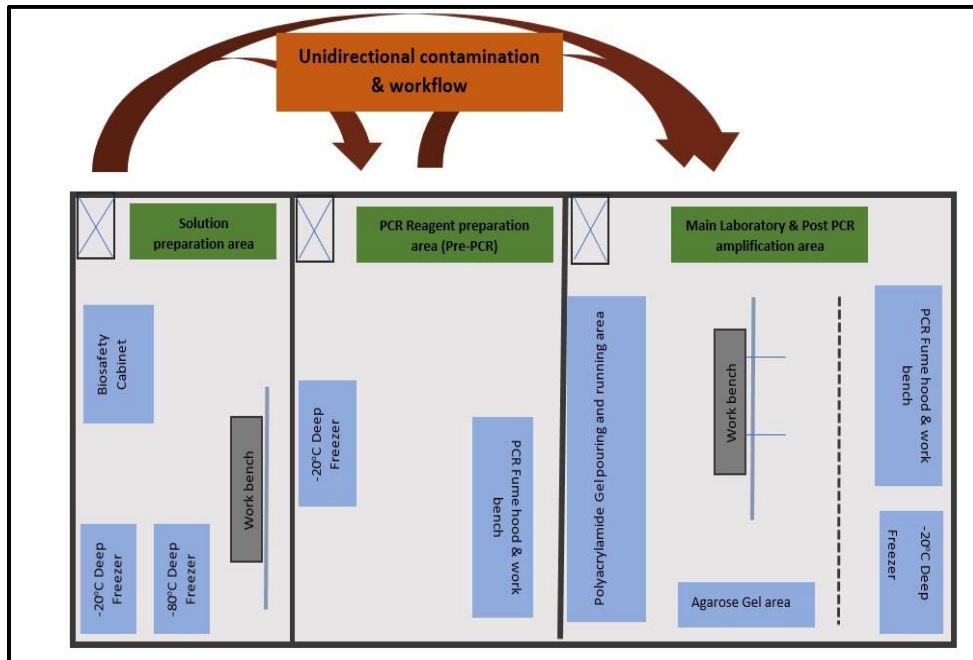
Given the facts above, maintaining and practicing cleanliness is necessary in many ways when establishing a contamination- and RNase-free laboratory. In this chapter, I plan to address both problems and describe some general guidelines for the establishment and maintenance of a clean environment for any DNA- and RNA- based research laboratory.

All laboratories have contamination problems. These problems are unavoidable without significant investment in construction and maintenance of clean room facilities. When establishing a contamination-free laboratory, there are a lot of factors needed to be considered. One essential concept we practice in our lab is maintaining two separate “Clean Room” areas besides the main laboratory. The cleaner of these two areas, we use as the stock solution preparation room. The other area is for PCR reagent preparation, which will be explained later in detail. This makes it is easier to maintain a unidirectional contamination flow across all three lab spaces (**Figure 31**). In general: whatever comes out

of our stock solution prep room, never goes back to that room again. It will either go to the PCR -Clean Room or to the main lab.

Figure 31: Illustration of different layers of a DNA and RNA research laboratory.

This shows the different layers of a research laboratory which conducts both PCR and RNA based experiments. Maintaining three separate lab spaces helps to establish a unidirectional contamination flow. Arrows depict how the reagents and other solutions must be carried in and out of each lab space.



1. PCR contamination prevention

PCR contamination prevention starts with the construction or set-up of a clean PCR laboratory. At a minimum, two areas should be designated for PCR testing: Pre- and Post-PCR (**Figure 31 & 32**). One room or area should be designated specifically for Pre-PCR. Maintaining a separate room for this is very critical. We maintain a PCR-Clean room as the pre-PCR area. This pre-PCR area contains storage space for the reagents needed for PCR. Stock- bought PCR related enzymes, DNA primers, dNTPs, salts, and buffer solutions are mostly stored in our PCR-Clean room. There we dilute or aliquot these reagents either into Falcon tubes or into microcentrifuge tubes. This helps to prevent the contamination of the entire stock solution. This room must be only used for PCR sample preparation. Stock bought DNA oligos are stored inside the clean room freezer and dilute inside the clean room fume hood. Once they are diluted we bring the diluted samples in to the main laboratory and either use immediately under the fume hood or store in a separate -20°C refrigerator. A second room, Post-PCR area should be established for post amplification steps and analysis. This room is physically separated from the pre-PCR room. The post-PCR room has a separate fume hood to set up PCRs, thermal cyclers for amplification and any instrumentation i.e. gel electrophoresis apparatus needed for post-PCR analysis. Both areas have their own instruments and equipment such as microcentrifuges, pipettes, pipette tips, gloves, microcentrifuge tubes, 10- and 50-mL tubes and lab wipes.

The workflow should continue in one direction only. This is beneficial in minimizing the chances for contamination to occur. No materials, supplies or equipment from the post-PCR room should be taken into the pre-PCR room. Similarly, nothing from the amplification and product detection area should be taken into the post-PCR room or the pre-PCR room. A PCR biocontainment hood is located inside the main laboratory. Once all the reagents necessary for the PCR is ready, they can be brought into the PCR hood.

a) Handling PCR materials

This PCR hood area has its own glove-boxes, pipettes, and pipette tips. An additional method we use for contamination control inside the PCR hood is the use of UV light, which is turned on for 30-60 minutes after each use prior to the next use. The use of gloves is essential; making sure to change gloves whenever you move from the pre-PCR area to post-PCR area.

b) Aerosol contamination

One way to minimize the contamination through aerosols is by briefly centrifuging the tubes prior to opening. If a PCR product is aerosolized, it will contain as many as 10^6 amplification products. If uncontrolled, within a relatively short time the buildup of aerosolized amplification products will contaminate laboratory reagents, equipment, and ventilation systems (Dieffenbach and Dreksler, 1993). To prevent this, the tubes should not be popped open, as this can create an aerosol.

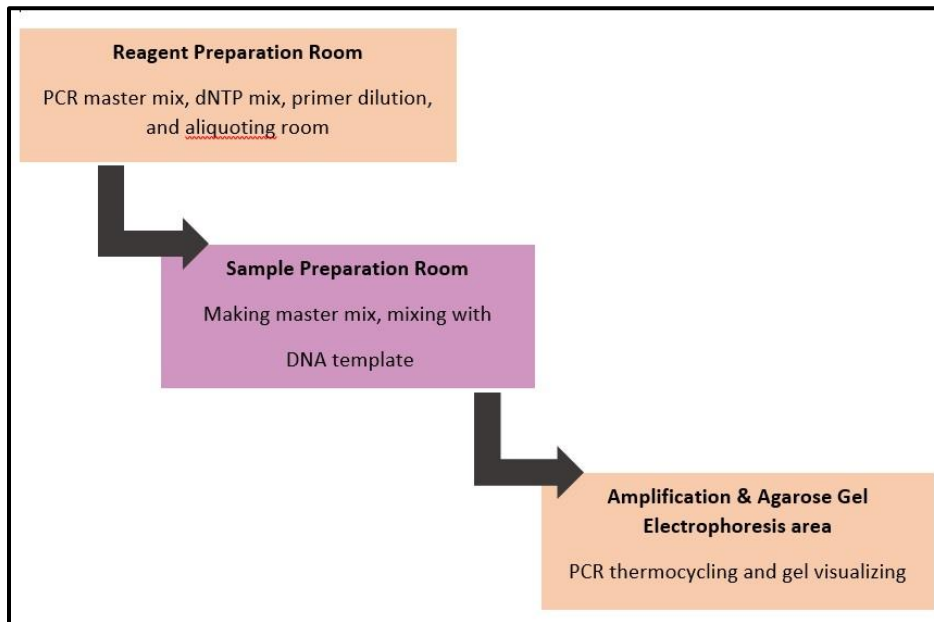
c) Sample and reagent preparation

Sample and reagent preparation areas play a major role when optimizing successful experiments. Quality of the reagents directly affects the experimental data. Therefore, it is necessary to make sure that these are contamination-free as well. Reagents that are used in sample preparation and analysis must be the highest quality available. Preparation and storing of the reaction buffers and enzymes should be done very carefully. For this, we maintain a separate -20°C freezer to store PCR reagents and mixtures, also store *Taq* polymerase stock in a separate -80°C freezer and take one tube at a time when it's necessary. In order to achieve a high-quality RNA or DNA, the purity of the stock solutions does matter a lot. Therefore, it is very important to focus on every single detail of possible ways to avoid contaminations along the preparation process.

2. RNA contamination prevention

As lab personal, it is necessary to acknowledge that contamination can happen at any time. Therefore, the most important thing is to take actions to minimize or prevent it. RNase contamination is always a concern when performing experiments involving RNA. RNase contamination can happen even with the cleanest of techniques and may have profound effects on experimental results. There are many ways that RNases can be introduced into the laboratory from bare hands to open windows and air vents, but the results are still the same: degraded RNA and, ultimately, compromised experiments.

Figure 32: Illustration of the work flow of a PCR laboratory. A unidirectional workflow helps to minimize PCR contaminations.



a) Handling RNA related materials

Although purified RNA may not contain RNases and could be perfectly intact, RNases can be introduced thereafter during experimental steps. When RNA contamination occurs, discarding the individual solutions that can be identified as contaminated is the most effective and least time-consuming way to eliminate the contamination. This is the easiest way to reassure that the samples or solutions have been contaminated. It is possible to get a lot of contaminations during RNA purification process as well. This is mainly through

the reused tubes, electrophoresis apparatuses, centrifuges, and common laboratory instruments and buffer solutions. We routinely use aerosol filter tips, aliquot all our solutions including small aliquots of water used for washing and usage, maintain different locations for RNA synthesis as well as RNA purification. To prevent contamination of the purified RNA samples, the container tubes must be opened and closed very carefully and must try to prevent the formation of aerosols. Tube-opening devices may be useful. Tips, tubes and all the material use in these procedures must be very clean and sterilized. Placement of microcentrifuge tubes in airtight self-sealing storage bags is an effective way of keeping them free from contamination. Wearing gloves and using Nuclease-free water, which can inhibit enzymatic reactions or Milli-Q™ (Millipore) purified water is good practice for most RNA work. If you are not sure that an aliquot of reagent is RNase free, discard it. RNase away the instruments including pipettes, and lab benches in a routinely way is a best way to minimize these contaminants. Sampling procedure is also another potential source of contamination. Containers which are used for sampling is a good way to add contaminations during sampling. Containers should be washed with suitable analyte-free detergent and rinsed with distilled water.

b) RNA gels

In an RNA laboratory, most of the experiments based on working with urea-polyacrylamide gels. Special precautions should be taken when making and handling these gels. In our lab, we maintain separate areas for gel pouring and gel electrophoresis. This helps to minimize the amount of contamination of the gels and the RNA which is ready to

be purified. Cleaning the glass plates and the electrophoresis apparatuses is very important as well.

c) Sample and reagent preparation

In our lab reaction buffers and gel reagents for RNA purification are prepared in the cleanest area or the solution preparation area (**Figure 27**). Once these stock solutions are made, we bring them to the main laboratory. During this transfer period, we make sure not to touch anything, even the door knobs or keys until the stock bottle is properly transferred and placed inside the main lab. Always wearing a fresh pair of gloves when handling freshly made stock buffer and other reagent solutions helps to minimize any cross contaminations during the transfer process. It is recommended to use MiliQ water for stock solution preparation. RNase contamination can be seen in both stock and personal laboratory reagents. These stock solutions are susceptible for contamination when they are kept for long periods of time.

d) Cleaning

The primary source of contamination in many laboratories is carryover from previous experiments. To avoid all the possible unwanted RNA and RNase contaminations, first we clean the glass plates from regular tap water and detergent, second, three times with double distilled water and finally another three more times with MiliQ water and 70% Ethanol. Periodically we clean with RNase Away, especially if there has been recent evidence of RNase activity (e.g., RNA smearing upon electrophoresis). Electrophoresis gel apparatuses

are also properly rinsed and dry-out prior to use. Most importantly, the gel running area should be cleaned and RNase away thoroughly, to get rid of any contaminations from previous experiments. We follow the same procedure for agarose gels as well. Using a new pair of gloves in between each step is very important, to avoid cross contaminations.

Discussion

Due to the significant improvements in every area of amplification technology it should be emphasized that PCR and other amplification techniques remain susceptible to carry-over contamination. False-positive findings can be identified not only in manually implemented research laboratory techniques but also in most of the commercially-available automated systems. Several precautions are required to be considered when designing a contamination- and RNase-free laboratory such that the laboratory is operated in a way that prevents cross contamination and RNase from previous reactions and assays, both of which can lead to false positive results. In a DNA and RNA involved laboratory, every precaution must be taken to ensure that results are accurate and unexpected degradations or bands do not occur on the gels. In this chapter, I tried to illustrate how to set up a contamination- and RNase- free laboratory. I have clearly discussed principles and methods to eliminate both DNA and RNA contaminants and offered tips to prevent contaminations in the future. Avoiding these types of contaminations can be done synergistically. By critically following these recommendations, anyone who works in an DNA-RNA laboratory can prevent RNase and possible PCR contaminations from ever becoming a problem in the experimental results. If standard sterile techniques that are generally applied to other molecular biology procedures are applied to PCR and RNA purification, then the risk of contamination will be greatly reduced. Most importantly, common sense is more required. The laboratory must go through a routine check-up to continuously monitor the many different types and sources of contamination. When contamination is found, active steps must be implemented to remove or minimize the contamination. In the absence of a proper

contaminant minimization and monitoring program, laboratory results are always at a risk of false or negative results.

References

- Adams, P.L., Stahley, M.R., Wang J., Strobel S.A., (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45-50.
- Attwater, J., Wochner, A., and Holliger., P., (2013). In-ice evolution of RNA polymerase ribozyme activity. *Nature Chemistry* **5**, 1011–1018.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A., (2000). The complete structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905-920.
- Bartel, D.P., and Szostak, J. W., (1993). Isolation of new ribozymes from a large pool of random sequence. *Science* **261**, 1411-1418.
- Becker, S., Schneider, C., Okamura, H., Crisp, A., Amatov, T., Dejmek, M., and Carell, T., (2018). Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nature Communications* **9**, 163-171.
- Been, M.D., Barfod, E T., Burke, J.M., Price, J.V., Tanner, N. K., Zaug, A. J., Cech, T.R., (1987). Structures involved in *Tetrahymena* rRNA self-splicing and RNA enzyme activity. *Cold Spring Harbor Symposia on Quantitative Biology* **52**,147-57.
- Been, M.D., and Cech, T.R., (1986). One binding site determines sequence specificity of *Tetrahymena* pre-rRNA self-splicing, *trans*-splicing, and RNA enzyme activity. *Cell* **47**, 207-216.
- Biebricher, C.K., and Luce, R., (1992). *In vitro* recombination and terminal elongation of RNA by Q beta replicase. *EMBO J.* **11**, 5129-5135.

- Briones, C., Stich, M., and Manrubia, S.C., (2009). The dawn of the RNA world: toward functional complexity through ligation of random RNA oligomers. *RNA* **15**, 743-749.
- Butlerow, A., (1861). Bildung einer zuckerartigen substanz durch synthese. *European Journal of Organic Chemistry* **120**, 295-298.
- Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'SSouza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., et al. (2002). The comparative RNA web (CRW) site: An online database of comparative sequence and structural information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics* **3**, <http://www.rna.icmb.utexas.edu/>
- Cech, T. R., (1988). Conserved sequences and structures of group I introns: building an active site for RNA catalysis. *Gene* **73**, 259-271.
- Cech, T.R., (1990). Self-splicing of group I introns. *Annu. Rev. Biochem* **59**, 543-568.
- Crick, F.H., (1968). The origin of the genetic code. *J.Mol. Biol.* **38**, 367-379.
- Da Silva, L., Maurel, M., Deamer, D., (2015). Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *Journal of Molecular Evolution* **80**, 86-97.
- Damer, D., Deamer, D., (2015). Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life*, **5**, 872-887.
- Da Silva, L., Maurel, M., Deamer, D., (2015). Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *Journal of Molecular Evolution* **80**, 86-97.

- Deamer, D., Singaram, S., Rajamani, S., Kompanichenko, V., and Guggenheim, S., (2006). Self-assembly processes in the prebiotic environment. *Philos Trans R Soc Lond B Biol Sci.*, **361**(1474), 1809–1818.
- DeGuzman, V., Vercoutere, W., Shenasa, H., Deamer, D., (2014). Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. *Journal of Molecular Evolution* **78**, 251-262.
- Delaye, L., and Lazcano, A., (2005). Prebiological evolution and the physics of the origin of life. *Phys. Life Rev.* **2**, 47-64.
- Dieffenbach, C.W., and Drexler, G.S., (1993). Setting up a PCR laboratory. *PCR Meth. Appl.* **3**, 2-7.
- Doudna, J.A., Cormack, B.P., and Szostak, J.W. (1989). RNA structure, not sequence, determines the 50 splice-site specificity of a group I intron. *Proc. Natl Acad. Sci. USA* **86**, 7402–7406.
- Doudna, J.A., and Szostak, J.W., (1989). RNA-catalyzed synthesis of complementary-strand RNA. *Nature* **339**, 519-522.
- Doudna, J.A., Couture, S., and Szostak, J.W., (1991). A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251**, 1605-1608.
- Doudna, J.A., and Cech, T.R., (1995). Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* **1**(1), 36-45.
- Draper, W.E., Hayden, E.J., and Lehman, N., (2008). Mechanisms of covalent self-assembly of the *Azoarcus* ribozyme from four fragment oligonucleotides. *Nucleic Acids Research* **36**, 520-531.

- Ferris, J.P., Hill Jrt, A. R., Uut, R., and Orgel, L. E., (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59-61.
- Ferris, J.P., and Ertem, G., (1993). Montmorillonite Catalysis of RNA Oligomer Formation in Aqueous Solution. A Model for the Prebiotic Formation of RNA. *J. Am. Chem. Soc.* **115**, 12270-12275.
- Gibard, C., Bhowmik, S., Karki, M., Kim, E., and Krishnamurthy, R., (2018). Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nature Chemistry* **10**, 212-217.
- Gilbert, W., (1986). Origin of life: The RNA world. *Nature* **319**, 618.
- Golden, B.L., Kim, H., and Chase E., (2005). Crystal structure of a phage *Twort* group I ribozyme product complex. *Nature Structural and Molecular Biology* **12**, 82-89.
- Guo, F., Gooding, A.R., Cech, T.R., (2004). Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Molecular Cell* **16**, 351-362.
- Hanna, R., and Doudna, J.A., (2000). Metal ions in ribozyme folding and catalysis. *Curr. Opin. Chem. Biol.* **4**, 166-170.
- Haugen. P., Simon, D.M., and Bhattacharya, D., (2005). The natural history of group I introns. *Trends in Genetics* **21**, 111-119.
- Hayden, E.J., and Lehman, N., (2006). Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem. Biol.* **13**, 909-918.
- Higgs, P.G., and Lehman, N., (2015). The RNA world: molecular cooperation at the origins of life. *Nature Reviews Genetics* **16**, 7-17.
- Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds (1990). *PCR Protocols A guide to Methods and Applications*. Academic Press, San Diego, CA.

- Jayathilaka, T., and Lehman, N., (2018). Spontaneous covalent self-assembly of the *Azoarcus* ribozyme from five fragments. *ChemBioChem* **19**, 1-5.
- Joyce, J.F., (2000). The antiquity of RNA-based evolution. *Nature* **418**, 214-221.
- Kawamura, K., (2004). Behavior of RNA under hydrothermal conditions and the origins of life. *Int. Jour. Astrobiol.* **3**, 301-309.
- Kim, D.E., and Joyce, G.F., (2004). Cross catalytic replication of an RNA ligase ribozyme. *Chem. Biol.* **11**, 1505-1512.
- Kazakov, S.A., Balatskaya, S.V., and Johnston, B.H., (2006). Ligation of the hairpin ribozyme in cis induced by freezing and dehydration. *RNA* **12**, 446-456.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R., (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147-157.
- Kuo, L.Y., Davidson, L.A., and Pico, S., (1999). Characterization of the *Azoarcus* ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme. *Biochim. Biophys. Acta.* **1489**, 281-292.
- Lazcano, A., (2007). Prebiotic evolution and the origin of life: is a system-level understanding feasible in Systems Biology. Ed. By Rigoutsos, I. *Genomics*, **1** Oxford University Press, New York, 5-78.
- McGinness, K.E., and Joyce, G.F., (2003). In search of an RNA replicase ribozyme. *Chem. Biol.* **10**, 5-14.
- McSwiggen, J.A. and Cech, T.R., (1989). Stereochemistry of RNA cleavage by the *Tetrahymena* ribozyme and evidence that the chemical step is not rate-limiting. *Science* **244**, 679-683.

- Michel, F., and Westhof, E., (1990). Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585-610.
- Miller, S.L., (1953). A production of amino acids under possible primitive Earth conditions. *Science* **117**, 528-529.
- Mutschler, H., Wochner, A., Holliger P., (2015). Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry* **7**, 502-508.
- Neveu, M., Kim, H.J., and Benner, S. A. (2013). The ‘strong’ RNA world hypothesis: fifty years old. *Astrobiology* **13**, 391–403.
- Orgel, L.E., (2004). Prebiotic chemistry and the origin of the RNA world. *Cri. Revw. Biochem. Mol. Biol.* **39**, 99-123.
- Oró, J., (1960). Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* **2**, 407-412.
- Palasingam, K., and Shaklee, P.N., (1992). Reversion of Q beta RNA phage mutants by homologous RNA recombination. *Journal of Virology* **66**, 2435-2442.
- Persing, D.H., (1991). Polymerase chain reaction: trenches to benches. *Journal of Clinical Microbiology* **29**, 1281-1285.
- Powner, M.W., Gerland, B., and Sutherland, J.D., (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242.
- Price, J.V., and Cech, T.R., (1998). Determinants of the 3' splice site for self-splicing of the *Tetrahymena* pre-rRNA. *Genes Dev.* **2**, 1439-1447.
- Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D., (2008). *Origins of Life and Evolution of Biospheres* **38**, 57-74.

- Rangan, P., Masquida, B., Westhof, E., and Woodson, S.A., (2003). Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc.Natl. Acad. Sci. USA* **100**, 1574–1579.
- Rangan, P., and Woodson, S., (2003). Structural requirement for Mg^{2+} binding in the group I intron core. *Journal of Molecular Biology* **329**, 229-238.
- Reinhold-Hurek, B., and Shub, D.A., (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**, 173–176.
- Rich, A. (1962). On the problems of evolution and biochemical information transfer. In horizons in biochemistry, edited by M. Kasha and B. Pullman, Academic Press, New York, 103–126.
- Riley, C.A., and Lehman, N., (2003). Generalized RNA-directed recombination of RNA. *Chem. Biol.* **10**, 1233-1243.
- Ross, D.S., Deamer, D., (2016). Cycling and the thermodynamics and kinetics of prebiotic polymer synthesis. *Life* **6**, 1-12.
- Scott, W.G., Murray, J.B., and Strobel, S.A., (1996). Capturing the structure of a catalytic RNA intermediate: The hammerhead ribozyme. *Science* **20**, 2065-2069.
- Shan, S., Yoshida, A., Sun, S., Piccirilli, J.A., and Herschlag, D., (1996). Three metal ions at the active site of the Tetrahymena group I ribozyme. *Proct. Natl. Acad. Sci. USA* **22**, 12299-12304.
- Sievers, D., and von Kiedrowski, G. (1994). Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224.
- Tanner, M., and Cech, T., (1996). Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium *Azoarcus*. *RNA N.Y.N.*, **2**, 74-83.

- Usher, D. A., (1977). Early chemical evolution of nucleic acids: a theoretical model. *Science* **196**, 311-313.
- Vaidya, N., Manapat, M.I., Chen, I., Xulvi-Bruner, R., Hayden, E., and Lehman, N., (2012). Spontaneous network formation among cooperative RNA replicators. *Nature* **491**, 72-77.
- Walker, S. E., and Lorsch, J., (2003). RNA purification–precipitation methods. *Methods Enzymol.* **530**, 337–343.
- Watson, J.D., Crick, F.H.C., (1953). Molecular structure of Nucleic Acids. *Nature* **171**, 737-738.
- Wöhler, F., (1828). Ueber künstliche bildung des harnstoffs (artificial formation of urea). *Annalen der Physik* **88**, 253-256.
- Woodson, S.A., Cech, T.R., (1989). Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. *Cell* **57**, 335-345.
- Woodson, S.A., (2005). Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**, 324-330.
- Woese, C.R., Durge, D.H., Saxinger, W.C., and Durge, S.A., (1966). The molecular basis for the genetic code. *Proc. Natl. Acad. Sci. USA.* **55**, 966-974.
- Zaug, A.J., McEvoy, M.M., and Cech, T.R., (1993). Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochemistry* **32**, 7946-7953.